



# **STIC Search Report**

## **Biotech-Chem Library**

**STIC Database Tracking Number: 192050**

**TO: Ruixiang Li**  
**Location: REM/4D75/4C70**  
**Art Unit: 1646**  
**Monday, June 12, 2006**  
**Case Serial Number: 10/600816**

**From: Dwayne Hill**  
**Location: Biotech-Chem Library**  
**REM-1A64**  
**Phone: (571) 272-8972**

**Dwayne.hill@uspto.gov**

### **Search Notes**

Examiner Li,

See attached results.

If you have any questions about this search feel free to contact me at any time.

Thank you for using STIC search services!

Dwayne Hill  
Technical Information Specialist  
STIC Biotech/Chem Library  
(571) 272-8972

**THIS PAGE BLANK (USPTO)**



STIC-Biotech/ChemLib

199050

From: Li, Ruixiang  
Sent: Monday, June 05, 2006 3:22 PM  
To: STIC-Biotech/ChemLib  
Subject: Sequence search of Application No.10/600,816

Please do a standard search on:

- (i). SEQ ID NO: 3 against amino acid databases;
- (ii). SEQ ID NOS: 2 and 30-33 against nucleic acid databases.

Thank you very much!

Ruixiang Li  
GAU 1646  
REM 4D75  
Mail Box 4C70  
(571) 272-0875

STIC-Biotech/ChemLib  
199050  
JUN 05 2006

\*\*\*\*\*

Searcher: \_\_\_\_\_  
Searcher Phone: \_\_\_\_\_  
Date Searcher Picked up: \_\_\_\_\_  
Date completed: \_\_\_\_\_  
Searcher Prep Time: \_\_\_\_\_  
Online Time: \_\_\_\_\_

\*\*\*\*\*

Type of Search  
NA# \_\_\_\_\_ AA# \_\_\_\_\_  
S/L: \_\_\_\_\_ Oligomer: \_\_\_\_\_  
Encode/Transl: \_\_\_\_\_  
Structure #: \_\_\_\_\_ Text: \_\_\_\_\_  
Inventor: \_\_\_\_\_ Litigation: \_\_\_\_\_

\*\*\*\*\*

Vendors and cost where applicable  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
QUESTEL/ORBIT: \_\_\_\_\_  
LEXIS/NEXIS: \_\_\_\_\_  
SEQUENCE SYSTEM: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (Specify): \_\_\_\_\_

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2006, 15:07:32 ; Search time 430 Seconds  
(without alignments)  
10687.077 Million cell updates/sec

Title: US-10-600-816-2

Perfect score: 2456

Sequence: 1 ataacagcatgaagtgccgt.....ggcagcaaaaaaaaaa 2456

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1403666 seqs, 935554401 residues

Total number of hits satisfying chosen parameters: 2807332

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patents NA:\*

- 1: /EMC\_Celerra\_SIDS3/ptodata/2/ina/1 COMB.seq:\*
- 2: /EMC\_Celerra\_SIDS3/ptodata/2/ina/5 COMB.seq:\*
- 3: /EMC\_Celerra\_SIDS3/ptodata/2/ina/6A COMB.seq:\*
- 4: /EMC\_Celerra\_SIDS3/ptodata/2/ina/6B COMB.seq:\*
- 5: /EMC\_Celerra\_SIDS3/ptodata/2/ina/7 COMB.seq:\*
- 6: /EMC\_Celerra\_SIDS3/ptodata/2/ina/H COMB.seq:\*
- 7: /EMC\_Celerra\_SIDS3/ptodata/2/ina/PCTUS COMB.seq:\*
- 8: /EMC\_Celerra\_SIDS3/ptodata/2/ina/PP COMB.seq:\*
- 9: /EMC\_Celerra\_SIDS3/ptodata/2/ina/RE COMB.seq:\*
- 10: /EMC\_Celerra\_SIDS3/ptodata/2/ina/backfiles.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2456	100.0	2456	3	US-09-949-016-572 Sequence 572, App
2	2289	93.2	2289	3	US-09-949-016-3613 Sequence 3613, App
3	1215.4	49.5	26086	3	US-09-949-016-15355 Sequence 15355, A
4	1215.4	49.5	26238	3	US-09-949-016-12314 Sequence 12314, A
5	1126	45.8	1212	3	US-09-188-930-249 Sequence 249, App
6	1126	45.8	1212	3	US-09-312-283C-249 Sequence 249, App
7	1031.6	42.0	1228	3	US-09-495-050A-223 Sequence 223, App
8	600.6	24.5	601	3	US-09-949-016-29173 Sequence 29173, A
9	600.6	24.5	601	3	US-09-949-016-29174 Sequence 29174, A
10	600.6	24.5	601	3	US-09-949-016-126957 Sequence 126957, A
11	600.6	24.5	601	3	US-09-949-016-126958 Sequence 126958, A
12	418.6	17.0	601	3	US-09-949-016-29170 Sequence 29170, A
13	418.6	17.0	601	3	US-09-949-016-126954 Sequence 126954, A
14	292.2	11.9	311	3	US-09-188-930-4 Sequence 4, Appli
15	292.2	11.9	311	3	US-09-312-283C-4 Sequence 4, Appli
16	204.4	8.3	206	4	US-09-880-107-295 Sequence 295, App
17	201.4	8.2	1880	3	US-09-620-312D-897 Sequence 897, App
18	179.4	7.3	2759	3	US-09-495-050A-291 Sequence 291, App
19	139.4	5.7	2484	3	US-09-276-531-46 Sequence 46, Appl
20	117.8	4.8	389504	3	US-09-949-016-11774 Sequence 11774, A
21	116.8	4.8	87039	3	US-09-949-016-15691 Sequence 15691, A
22	115.6	4.7	75929	3	US-09-949-016-15543 Sequence 15543, A
23	115.6	4.7	75929	3	US-09-949-016-15544 Sequence 15544, A

ALIGNMENTS

RESULT 1

US-09-949-016-572

; Sequence 572, Application US/09949016

; Patent No. 6812339

; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

; FILE REFERENCE: CL001307

; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 2000-04-14

; PRIOR APPLICATION NUMBER: 60/241,755

; PRIOR FILING DATE: 2000-10-20

; PRIOR APPLICATION NUMBER: 60/237,768

; PRIOR FILING DATE: 2000-10-03

; PRIOR APPLICATION NUMBER: 60/231,498

; PRIOR FILING DATE: 2000-09-08

; NUMBER OF SEQ ID NOS: 207012

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 572

; LENGTH: 2456

; TYPE: DNA

; ORGANISM: Human

US-09-949-016-572

Query Match 100.0%; Score 2456; DB 3; Length 2456;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATAACAGCATGAAGTCCCTGGAACTGGATAGGCGGTCTCTCCCTCGACCTCCCC 60

DB 1 ATAACAGCATGAAGTCCCTGGAACTGGATAGGCGGTCTCTCCCTCGACCTCCCC 60

QY 61 TCCTTCTCCCTCTGCTCACCTCGCTCGTTCCTCCCGGAGGCGCGCTTTATA 120

DB 61 TCCTTCTCCCTCTGCTCACCTCGCTCGTTCCTCCCGGAGGCGCGCTTTATA 120

QY 121 ACAACTGCTCAGAGTCGAGGCGGGATAGTGTCTCAAGTCTCCCGACCTGAGGAG 180

DB 121 ACAACTGCTCAGAGTCGAGGCGGGATAGTGTCTCAAGTCTCCCGACCTGAGGAG 180

QY 181 CTCGCTGTGCTCCCTTTGGCGGGAGCAGCACAAGTTCACGGCCACCGCTTGGC 240

DB 181 CTCGCTGTGCTCCCTTTGGCGGGAGCAGCACAAGTTCACGGCCACCGCTTGGC 240

QY 241 ACTAGGGTCCAGATGGCTACAACAGTCCCTGTGTTGCCCAATGGCGCTGAATCCAA 300

Sequence 15945, A

Sequence 15944, A

Sequence 14934, A

Sequence 14194, A

Sequence 12214, A

Sequence 47244, A

Sequence 13122, A

Sequence 47245, A

Sequence 48485, A

Sequence 13175, A

Sequence 11886, A

Sequence 17213, A

Sequence 12503, A

Sequence 14263, A

Sequence 12287, A

Sequence 14864, A

Sequence 88852, A

Sequence 88853, A

Sequence 204419, A

Sequence 204420, A

Sequence 204489, A

|||||  
Db 241 ACTAGGGTCCAGATGGCTACACAGTCCCTGATGGTTGCCGAATGGCCCTGAAATCCAA 300  
QY 301 GTACTACAGACTTTGTGATAAGCTTGAAGCTTGGGGCATCGCTTAGAAGACGGTGGCCAC 360  
Db 301 GTACTACAGACTTTGTGATAAGCTTGAAGCTTGGGGCATCGCTTAGAAGACGGTGGCCAC 360  
QY 361 AGCCGGGTTGTGACCTCGGTGGCTTCATGTCTCACTCTCCGATCCTCGTCTGCAAGGT 420  
Db 361 AGCCGGGTTGTGACCTCGGTGGCTTCATGTCTCACTCTCCGATCCTCGTCTGCAAGGT 420  
QY 421 GCAGGACTCCACAGCGGAAATCGTCCCTACTCAGTTTCTCTTCCTCCTCGGTGTGT 480  
Db 421 GCAGGACTCCACAGCGGAAATCGTCCCTACTCAGTTTCTCTTCCTCCTCGGTGTGT 480  
QY 481 GGGCATCTTTGGCTCACCTTCGCCCTTCACTCGGACTGGACGGGAGCACAGGGCCAC 540  
Db 481 GGGCATCTTTGGCTCACCTTCGCCCTTCACTCGGACTGGACGGGAGCACAGGGCCAC 540  
QY 541 AGCGTCTCTCTTTGGGATCCTCTTTTCCATCTGCTTCTCTGCTGCTGCTGCTCATGC 600  
Db 541 AGCGTCTCTCTTTGGGATCCTCTTTTCCATCTGCTTCTCTGCTGCTGCTGCTCATGC 600  
QY 601 TGTCACTGACCAAGCTCGTGGGGGAGGAGCCCTTCCCTGTTGGTCAATCTGGG 660  
Db 601 TGTCACTGACCAAGCTCGTGGGGGAGGAGCCCTTCCCTGTTGGTCAATCTGGG 660  
QY 661 TCTGGCGTGGGCTTCAGCTAGTCAGGATGTTATCGCTATTGAATATATTGTCCTGAC 720  
Db 661 TCTGGCGTGGGCTTCAGCTAGTCAGGATGTTATCGCTATTGAATATATTGTCCTGAC 720  
QY 721 CATGAATAGGACCAAGCTCAATGTCTTTTCTGAGCTTCCGCTCCTCGTGGCAATGAAGA 780  
Db 721 CATGAATAGGACCAAGCTCAATGTCTTTTCTGAGCTTCCGCTCCTCGTGGCAATGAAGA 780  
QY 781 CTTTGTCTCTCTGCTCACCTAGTCGTCTCTCTTGTATGGGCTGACCTTCTCATGCTCTC 840  
Db 781 CTTTGTCTCTCTGCTCACCTAGTCGTCTCTCTTGTATGGGCTGACCTTCTCATGCTCTC 840  
QY 841 CTTTCACTTCTGTGGTTCCTTCAAGGGCTGGAAGAGACATGGGGCCACATCTACCTCAC 900  
Db 841 CTTTCACTTCTGTGGTTCCTTCAAGGGCTGGAAGAGACATGGGGCCACATCTACCTCAC 900  
QY 901 GATGCTCTCTCCATTTGCCATCTGGGTGGCTGGATCACCTGCTCATGCTTCTTGACTT 960  
Db 901 GATGCTCTCTCCATTTGCCATCTGGGTGGCTGGATCACCTGCTCATGCTTCTTGACTT 960  
QY 961 TGACCGCAGGTGGATGACACCATCTCAGCTCCGCTTGGCTGGCAATGGGCTGGGTGT 1020  
Db 961 TGACCGCAGGTGGATGACACCATCTCAGCTCCGCTTGGCTGGCAATGGGCTGGGTGT 1020  
QY 1021 CTTGTGGCTTATGTTAGTCCCGAGTTTGGGTGCTCACAAGACCAACCAACCCCATGGA 1080  
Db 1021 CTTGTGGCTTATGTTAGTCCCGAGTTTGGGTGCTCACAAGACCAACCAACCCCATGGA 1080  
QY 1081 TTATCTGTTGAGGATGCTTTCTGTAAACCTCAACTCGTGAAGAGAGCTATGGTGTGA 1140  
Db 1081 TTATCTGTTGAGGATGCTTTCTGTAAACCTCAACTCGTGAAGAGAGCTATGGTGTGA 1140  
QY 1141 GAACAGAGCTTACTCTCAGAGGAAATCACTCAAGGTTTGAAGACAGGAGCACGCT 1200  
Db 1141 GAACAGAGCTTACTCTCAGAGGAAATCACTCAAGGTTTGAAGACAGGAGCACGCT 1200  
QY 1201 CTATGCCCCCTATTCCACATTTTTCAGCTGAGAACACGCTCCCCAAAGGAATTTCTC 1260  
Db 1201 CTATGCCCCCTATTCCACATTTTTCAGCTGAGAACACGCTCCCCAAAGGAATTTCTC 1260  
QY 1261 CATCCCAACGGGCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAGAGAGGG 1320  
Db 1261 CATCCCAACGGGCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAGAGAGGG 1320  
QY 1321 CAGCTAACTCTGTCTGAAGAGTGGGACAAATGACCGGGCGGAGATCTAGCGGAGC 1380  
Db 1321 CAGCTAACTCTGTCTGAAGAGTGGGACAAATGACCGGGCGGAGATCTAGCGGAGC 1380

1381 TCAAGAGGATGTGGCGAAATCTTGAGTCTTCTGAGAAAACTGTACAAGACACTACGGGA 1440  
1381 TCAAGAGGATGTGGCGAAATCTTGAGTCTTCTGAGAAAACTGTACAAGACACTACGGGA 1440  
1441 ACAGTTTGGCTCCCTCCAGCCTCAACCAAAATCTTCCATGCTGGGCTGATGGGCT 1500  
1441 ACAGTTTGGCTCCCTCCAGCCTCAACCAAAATCTTCCATGCTGGGCTGATGGGCT 1500  
1501 AGTAAGACTCCAGTTCTTAGAGCGCTGTAGTATATTTTTTTTTTTTGTCTCATCTTTGG 1560  
1501 AGTAAGACTCCAGTTCTTAGAGCGCTGTAGTATATTTTTTTTTTTTGTCTCATCTTTGG 1560  
1561 ATACTTCTTTTAAAGTGGGAGTCTCAGGCAATCAAGTTTAGACCCCTTACTCTTTTGT 1620  
1561 ATACTTCTTTTAAAGTGGGAGTCTCAGGCAATCAAGTTTAGACCCCTTACTCTTTTGT 1620  
1621 GTTTTTTGAACAGGATCTTGTCTGTCAACCAAGCTTGAAGTGGATCACAG 1680  
1621 GTTTTTTGAACAGGATCTTGTCTGTCAACCAAGCTTGAAGTGGATCACAG 1680  
1681 CCAGTGCAGCTCGACCACTGTCTCAAGCAATCCTCCCATCTCCATCTCCAAAGTG 1740  
1681 CCAGTGCAGCTCGACCACTGTCTCAAGCAATCCTCCCATCTCCATCTCCAAAGTG 1740  
1741 CTGGGATGACAGCGCTGAGCCACAGCTCCAGCCCTAGGCGCTTAAATCTTGTGTTATTT 1800  
1741 CTGGGATGACAGCGCTGAGCCACAGCTCCAGCCCTAGGCGCTTAAATCTTGTGTTATTT 1800  
1801 CCATGACTAAAGTCTTGTCTGCTCAGCTCAGCTGGCTCACACAGCTCTAGGGCGCTG 1860  
1801 CCATGACTAAAGTCTTGTCTGCTCAGCTCAGCTGGCTCACACAGCTCTAGGGCGCTG 1860  
1861 CTCCTTAACTCACAAGTGGGTTTGTGAGGCTCTGTGGCCAGAGCAGACCTGCATATCT 1920  
1861 CTCCTTAACTCACAAGTGGGTTTGTGAGGCTCTGTGGCCAGAGCAGACCTGCATATCT 1920  
1921 GAGCAAAAATAGCAAAAGCCTCTCTCAGCCCACTGGCCTGAATCTACACTGGAAGCAAC 1980  
1921 GAGCAAAAATAGCAAAAGCCTCTCTCAGCCCACTGGCCTGAATCTACACTGGAAGCAAC 1980  
1981 TTGCTGGCACCCCGCTCCCAACCTTCTTCCCTGGGTAGGAGGCTTAAAGATCACCC 2040  
1981 TTGCTGGCACCCCGCTCCCAACCTTCTTCCCTGGGTAGGAGGCTTAAAGATCACCC 2040  
2041 TAAATTTACTCATCTCTTAGTGTCTCAGCTCAGATTGGGCTCAGCAGCTCCCAGACCA 2100  
2041 TAAATTTACTCATCTCTTAGTGTCTCAGCTCAGATTGGGCTCAGCAGCTCCCAGACCA 2100  
2101 ATTCAAGTCAACCCCTCTCTTCTTGGCACTGTCCCAAACTTGTCTCAATTCGAGATC 2160  
2101 ATTCAAGTCAACCCCTCTCTTCTTGGCACTGTCCCAAACTTGTCTCAATTCGAGATC 2160  
2161 TAAATCTCCCTTACGCTCTGCGAGGAAATCTTTTACCGCAATTTACCGCTGCAATTCGTA 2220  
2161 TAAATCTCCCTTACGCTCTGCGAGGAAATCTTTTACCGCAATTTACCGCTGCAATTCGTA 2220  
2221 CTCCTTGTGAGGAAATTTGTAGATCATTTCTCACTTCAAAATTCCTGGGCTGATCTTCT 2280  
2221 CTCCTTGTGAGGAAATTTGTAGATCATTTCTCACTTCAAAATTCCTGGGCTGATCTTCT 2280  
2281 CTCATCTTGCAACCCAACTCTCTGTAATAGATTTACCGCAATTTACCGCTGCAATTCGTA 2340  
2281 CTCATCTTGCAACCCAACTCTCTGTAATAGATTTACCGCAATTTACCGCTGCAATTCGTA 2340  
2341 GTGGCATGGTCTCTTAATGGAGGAGTGTTCATTTATTAAGTTATTTCACCTGAGTAT 2400  
2341 GTGGCATGGTCTCTTAATGGAGGAGTGTTCATTTATTAAGTTATTTCACCTGAGTAT 2400  
2401 GCAATAAAGATGTGGTGGCCACTCTTTTATGTTGGTGGCAGCAAAAAA 2456  
2401 GCAATAAAGATGTGGTGGCCACTCTTTTATGTTGGTGGCAGCAAAAAA 2456

## RESULT 2

US-09-949-016-3613  
; Sequence 3613, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3613  
; LENGTH: 2289  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-3613

Query Match 93.2%; Score 2289; DB 3; Length 2289;

Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2289; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	155	CCAAGTCTCCCCAGACACTGAGGAGCTCGCCTGCTGCCCTTTGGCGCGGGAAGCAGC	214
DB	1	CCNAGGTCTCCCCAGACACTGAGGAGCTCGCCTGCTGCCCTTTGGCGCGGGAAGCAGC	60
QY	215	ACCAAGTCTCAGCGCCCAACGCTTGGCACTAGGCTCCAGATGCTACAAAGTCCCTGAT	274
DB	61	ACCAAGTCTCAGCGCCCAACGCTTGGCACTAGGCTCCAGATGCTACAAAGTCCCTGAT	120
QY	275	GGTTGGCGCAATGGCTCGAAATCCAAAGTACTACAGACTTTGTGATTAAGCTGAAGCTGG	334
DB	121	GGTTGGCGCAATGGCTCGAAATCCAAAGTACTACAGACTTTGTGATTAAGCTGAAGCTGG	180
QY	335	GGCATCGTCTAGAAACGGTGGCCACAGCGGGGTTGTGACCTCGGTGGCTTCATGCTC	394
DB	181	GGCATCGTCTAGAAACGGTGGCCACAGCGGGGTTGTGACCTCGGTGGCTTCATGCTC	240
QY	395	ACTCTCCGATCCTCGTCTGCAAGGTGCAGGACTCCAAACAGGCGGAAATGCTGCCCTACT	454
DB	241	ACTCTCCGATCCTCGTCTGCAAGGTGCAGGACTCCAAACAGGCGGAAATGCTGCCCTACT	300
QY	455	CAGTTTCTCTCTCGTGGGTGTGGGATCTTTGGGCTCACCTTGGCTTCATCATC	514
DB	301	CAGTTTCTCTCTCGTGGGTGTGGGATCTTTGGGCTCACCTTGGCTTCATCATC	360
QY	515	GGACTGCGGGAGACAGCGGCCACACGCTTCTCTCTTTGGGATCCTCTTTCCATC	574
DB	361	GGACTGCGGGAGACAGCGGCCACACGCTTCTCTCTTTGGGATCCTCTTTCCATC	420
QY	575	TGCTTCTCTCTCGTGGGTCTAGTGTGAGTCAAGCTCGTCCGGGGGAGGAAG	634
DB	421	TGCTTCTCTCTCGTGGGTCTAGTGTGAGTCAAGCTCGTCCGGGGGAGGAAG	480
QY	635	CCCTTTCCCTGTGGTGTGGGTCTGGCGGTTCAGCTAGTCCAGGATGT	694
DB	481	CCCTTTCCCTGTGGTGTGGGTCTGGCGGTTCAGCTAGTCCAGGATGT	540
QY	695	ATCGCTATTGAATATATGCTGACCAATAGGACCAAGCTCAATGCTTTCTGAG	754
DB	541	ATCGCTATTGAATATATGCTGACCAATAGGACCAAGCTCAATGCTTTCTGAG	600
QY	755	CTTTCGGCTCCTCGTCCCAATGAAGACTTTGTCTCTGCTACCTACGCTCTCTCTTG	814
DB			

DB	601	CTTTCCGCTCCTCGTCGCAATGAAGACTTTGTCTCTCTGCTCACCTACGTCTCTCTTG	660
QY	815	ATGGCGCTGACCTTCTCTCATGTCTCTCTTCACTTCTGTGGTCTCTTACGGGCTGAAG	874
DB	661	ATGGCGCTGACCTTCTCTCATGTCTCTCTTCACTTCTGTGGTCTCTTACGGGCTGAAG	720
QY	875	AGACATGGGGGCCACACATCTACCTCAGATGCTCTCTTCCATTGCCATCTGGTGGCTGG	934
DB	721	AGACATGGGGGCCACACATCTACCTCAGATGCTCTCTTCCATTGCCATCTGGTGGCTGG	780
QY	935	ATCACCTCTGCTCATGTCTCTCTGACTTTGACCGGAGTGGGATGACACCATCTCAGCTCC	994
DB	781	ATCACCTCTGCTCATGTCTCTCTGACTTTGACCGGAGTGGGATGACACCATCTCAGCTCC	840
QY	995	GCCTTGGCTGCCAATGGCTGGGTGTTCTGTGGCTTATGTTAGTCCCGAGTCTTGGCTG	1054
DB	841	GCCTTGGCTGCCAATGGCTGGGTGTTCTGTGGCTTATGTTAGTCCCGAGTCTTGGCTG	900
QY	1055	CTCAGAAAGCAAGCAAAACCCCATGGATTATCTGTGTGAGGATGCTTCTCTAAACCTCAA	1114
DB	901	CTCAGAAAGCAAGCAAAACCCCATGGATTATCTGTGAGGATGCTTCTCTAAACCTCAA	960
QY	1115	CTCGTGAAGAAGAGCTATGTTGTGGAGAACAGAGCTTCTCTCAAGAGGAAATCACTCAA	1174
DB	961	CTCGTGAAGAAGAGCTATGTTGTGGAGAACAGAGCTTCTCTCAAGAGGAAATCACTCAA	1020
QY	1175	GGTTTGAAGAGACAGGGACAGCTCTATGCCCTTATTCACACATTTTCAGCTGCAG	1234
DB	1021	GGTTTGAAGAGACAGGGACAGCTCTATGCCCTTATTCACACATTTTCAGCTGCAG	1080
QY	1235	AACCGCTCTCCCAAAAGGAATTTCTCCATCCACGGGCCACGCTTGGCGGAGCCCTTAC	1294
DB	1081	AACCGCTCTCCCAAAAGGAATTTCTCCATCCACGGGCCACGCTTGGCGGAGCCCTTAC	1140
QY	1295	AAAGACTATGAAGTAAAGAAAGAGGCGAGCTAACTCTGTCTCTGAAGAGTGGGACAAATGC	1354
DB	1141	AAAGACTATGAAGTAAAGAAAGAGGCGAGCTAACTCTGTCTCTGAAGAGTGGGACAAATGC	1200
QY	1355	AGCGGGCGGCGAGATCTAGGGGAGCTCAAGGGATGTGGCGGAAATCTTGAAGTCTCTG	1414
DB	1201	AGCGGGCGGCGAGATCTAGGGGAGCTCAAGGGATGTGGCGGAAATCTTGAAGTCTCTG	1260
QY	1415	AGAAACTGTACAGACACTACCGGAAACAGTTTGTCTCTCCCTCCAGCTCAACACCAATTT	1474
DB	1261	AGAAACTGTACAGACACTACCGGAAACAGTTTGTCTCTCCCTCCAGCTCAACCAATTT	1320
QY	1475	CTTCCATGCTGGGCTGATGTGGCTAGTAAGACTCCAGTTCTTAGAGGCGCTGTAGTAT	1534
DB	1321	CTTCCATGCTGGGCTGATGTGGCTAGTAAGACTCCAGTTCTTAGAGGCGCTGTAGTAT	1380
QY	1535	TTTTTTTTTTTGTCTCATCTTTGGATACTTTTTTAAGTGGAGTCTCAGGCAACTCA	1594
DB	1381	TTTTTTTTTTTGTCTCATCTTTGGATACTTTTTTAAGTGGAGTCTCAGGCAACTCA	1440
QY	1595	AGTTTACAGCTTACTCTTTTGTGTTTGTGTTTGAAGAGGATCTGTCTGTCCAGG	1654
DB	1441	AGTTTACAGCTTACTCTTTTGTGTTTGTGTTTGAAGAGGATCTGTCTGTCCAGG	1500
QY	1655	GCTTGAAGTCAAGTGTGCGATCACGCCAGTGCAGCTCGACCACTGTGCTCAAGCAA	1714
DB	1501	GCTTGAAGTCAAGTGTGCGATCACGCCAGTGCAGCTCGACCACTGTGCTCAAGCAA	1560
QY	1715	TCTCTCCATCTCCATCTCCAAAGTCTGGGATGACAGGGGTGAGCCACAGCTCCAGCC	1774
DB	1561	TCTCTCCATCTCCATCTCCAAAGTCTGGGATGACAGGGGTGAGCCACAGCTCCAGCC	1620
QY	1775	TAGGCGCTTAATCTGCTGTTATTTTCCATGGAAGTCTGAGTCTGAGTCTCAGC	1834
DB	1621	TAGGCGCTTAATCTGCTGTTATTTTCCATGGAAGTCTGAGTCTGAGTCTCAGC	1680
QY	1835	CTGGCTCACACAGCTCTAGGGGCTGCTCTCTTAACTCAAGTGGGTTTGTGAGGCTCT	1894
DB	1681	CTGGCTCACACAGCTCTAGGGGCTGCTCTCTTAACTCAAGTGGGTTTGTGAGGCTCT	1740



	Y	CGAGGAGTCTTCATTTGGTATTAAACGATGTGGTG	2418
	D	TGAAGGAGTCTTCATTTGGTATTAAACGATGTGGTG	2419
	B	CCACTCTTTTCATGTGGTGGCAGCA	2443
	Y	CGAGGAGTCTTCATTTGGTATTAAACGATGTGGTG	2460
	D	TGAAGGAGTCTTCATTTGGTATTAAACGATGTGGTG	2461
	B	CCACTCTTTTCATGTGGTGGCAGCA	2485

## RESULT 4

```

US-09-949-016-12314
; Sequence 12314, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12314
; LENGTH: 26238
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12314

```

Db	23495	ACCTGTGCTCAAGCAATCTCTCCCATCTCCCATCTCCCAAAAGTGCTGGGATGACAGCGGTGA	23554
Qy	1759	GCACAGCTCCCAAGCCTTAGCCCTTAATCTTGCTGTATTTTCCATGGACTAAAGGTC	1818
Db	23555	GCACAGCTCCCAAGCCTTAGCCCTTAATCTTGCTGTATTTTCCATGGACTAAAGGTC	23614
Qy	1819	GTCATCTGAGCTCACGCTGGCTCACACAGCTCTAGGGGCTGCTCTCTAACTCAAGTG	1878
Db	23615	GTCATCTGAGCTCACGCTGGCTCACACAGCTCTAGGGGCTGCTCTCTCTAACTCAAGTG	23674
Qy	1879	GGTTTTGTAGGCTCTGTGGCCAGACAGACCTGCATATCTGAGCAAAAATAGCAAAAG	1938
Db	23675	GGTTTTGTAGGCTCTGTGGCCAGACAGACCTGCATATCTGAGCAAAAATAGCAAAAG	23734
Qy	1939	CCTCTCTCAGCCACCTGGCTGAACTCACACTTGAAGCCAACTTGCTGGCACCCCGCTC	1998
Db	23735	CCTCTCTCAGCCACCTGGCTGAACTCTACACTTGAAAGCCAACTTGCTGGCACCCCGCTC	23794
Qy	1999	CCCAACCTTCTTGGCTGGGTAGGAGAGCTAAAGATCACCCCTAAATTTACTCATCTCTC	2058
Db	23795	CCCAACCTTCTTGGCTGGGTAGGAGAGCTAAAGATCACCCCTAAATTTACTCATCTCTC	23854
Qy	2059	TAGTGTGCTCTCACTTTGGGCTCAGCAGCTCCAGACACCAATTCACAGGTCACCCCTC	2118
Db	23855	TAGTGTGCTCTCACTTTGGGCTCAGCAGCTCCAGACACCAATTCACAGGTCACCCCTC	23914
Qy	2119	TCCTTCTGCACTCTCCCAAACTTGCTGTCAATTCGGAGATCTAATCTCCCCCTACGCTC	2178
Db	23915	TCCTTCTGCACTCTCCCAAACTTGCTGTCAATTCGGAGATCTAATCTCCCCCTACGCTC	23974
Qy	2179	TGCCAGGAATCTTTTTCAGACCTCAGTCAGCAAGCCCGGTGCTCTTGTTCAGAGAAAT	2238
Db	23975	TGCCAGGAATCTTTTTCAGACCTCAGTCAGCAAGCCCGGTGCTCTTGTTCAGAGAAAT	24034
Qy	2239	TGTAGATCATCTCACTTCAAAATCTCGGGGTGATATCTTCTCTCATCTTGCACCCCAAC	2298
Db	24035	TGTAGATCATCTCACTTCAAAATCTCGGGGTGATATCTTCTCTCATCTTGCACCCCAAC	24094
Qy	2299	CTCTGTAAATAGATTTTACCGCATTTACCGCTGCATTTCTGTAAGTGGGATGTGCTCTTAA	2358
Db	24095	CTCTGTAAATAGATTTTACCGCATTTACCGCTGCATTTCTGTAAGTGGGATGTGCTCTTAA	24154
Qy	2359	TGGAGGAGTGTTCATTTGTATATAAGTATTACCTCTGAGTATGCAATAAAGATGTGGTG	2418
Db	24155	TGGAGGAGTGTTCATTTGTATATAAGTATTACCTCTGAGTATGCAATAAAGATGTGGTG	24214
Qy	2419	CCACTCTTTCAVGGTGGCAGCA	2443
Db	24215	CCACTCTTTCAVGGTGGCAGCA	24239

## RESULT 5

```

RESULT 5
US-09-188-930-249
; Sequence 249, Application US/09188930A
; Patent NO. 6150502
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murlison, James Greg
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; TITLE OF INVENTION: and Methods For Their Use
; FILE REFERENCE: 11000.1011c1
; CURRENT APPLICATION NUMBER: US/09/188,930A
; CURRENT FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 249
; LENGTH: 1212
; TYPE: DNA
; ORGANISM: Human
US-09-188-930-249

```

Query Match		45.8%;	Score 1126;	DB 3;	Length 1212;	
Best Local Similarity		99.1%;	Pred. No. 2.8e-305;			
Matches 1132;		Conservative 0;	Mismatches 10;	Indels 0;	Gaps 0;	
QY	297	CCAAGTACTACAGACTTTGTGATAAGCTTGAAGCTTGGGGCATCGTCTAGAAAACGGTGG	356			
DB	16	CGAGGTACTACAGACTTTGTGATAAGCTTGAAGCTTGGGGCATCGTCTAGAAAACGGTGG	75			
QY	357	CCACAGCGGGGTGTGACCTCGGTGGCTTCATGCTCACTCCCGATCCTCGTCTGCA	416			
DB	76	CCACAGCTGGGTGTGACCTCGGTGGCTTCATGCTCACTCCCGATCCTCGTCTGCA	135			
QY	417	AGGTGAGGACTCCAAAGCGGAAAAATGCTGCCTACTCAGTTTCTTCTCTCTGGGTG	476			
DB	136	AGGTGAGGACTCCAAAGCGGAAAAATGCTGCCTACTCAGTTTCTTCTCTCTGGGTG	195			
QY	477	TGTTGGGCATCTTGGCCCTCAGCTTCGCTTTCATCATCGGACTGGACGGGACACAGGC	536			
DB	196	TGTTGGGCATCTTGGCCCTCAGCTTCGCTTTCATCATCGGACTGGACGGGACACAGGC	255			
QY	537	CCACAGCTTCTTCCCTTTGGATCCTCTTTTCCATCTGCTTCTCCTCGCTGGCTC	596			
DB	256	CCACAGCTTCTTCCCTTTGGATCCTCTTTTCCATCTGCTTCTCCTCGCTGGCTC	315			
QY	597	ATGCTGTCACTGACCAAGCTCGTCCGGGGAGGAGCCCTTTCCCTGTTGGTGATTC	656			
DB	316	ATGCTGTCACTGACCAAGCTCGTCCGGGGAGGAGCCCTTTCCCTGTTGGTGATTC	375			
QY	657	TGGGTCTGGCCGTGAGCTTACAGGTAGTCCAGGATGTTATCGCTATTGAATATATTGTC	716			
DB	376	TGGGTCTGGCCGTGAGCTTACAGGTAGTCCAGGATGTTATCGCTATTGAATATATTGTC	435			
QY	717	TGACCATGAATAGGACCAACGTCAATGTCTTTTCTGAGCTTTCCGCTCTCGTGGCAATG	776			
DB	436	TGACCATGAATAGGACCAACGTCAATGTCTTTTCTGAGCTTTCCGCTCTCGTGGCAATG	495			
QY	777	AAGACTTTGCTCCTGCTCAGCTACGTCCTCTTTGTGATGGCTGACCTTCTCTCATGT	836			
DB	496	AAGACTTTGCTCCTGCTCAGCTACGTCCTCTTTGTGATGGCTGACCTTCTCTCATGT	555			
QY	837	CTCTCTTACCTTCTGTGCTTCTTACCGGCTGGAGAGATGGGGCCACATCTACC	896			
DB	556	CTCTCTTACCTTCTGTGCTTCTTACCGGCTGGAGAGATGGGGCCACATCTACC	615			
QY	897	TACGATGCTCCTCTCCATTGGCATCTGGTGGCTGGATCACCTCTCATGCTTCCCTG	956			
DB	616	TACGATGCTCCTCTCCATTGGCATCTGGTGGCTGGATCACCTCTCATGCTTCCCTG	675			
QY	957	ACTTTGACCGCAGGTGGATGACACCATCTCAGCTCCGCTTGGCTGCAATGGCTGGG	1016			
DB	676	ACTTTGACCGCAGGTGGATGACACCATCTCAGCTCCGCTTGGCTGCAATGGCTGGG	735			
QY	1017	TGTTCTGTGGCTTATGTAGTCCGAGTGTGGCTGCTCAAAAGCAACGAAACCCCA	1076			
DB	736	TGTTCTGTGGCTTATGTAGTCCGAGTGTGGCTGCTCAAAAGCAACGAAACCCCA	795			
QY	1077	TGGATTATCCTGTGAGGATGCTTTCTGTAACCTCAACTCGTGAAGAGAGCTATGGTG	1136			
DB	796	TGGATTATCCTGTGAGGATGCTTTCTGTAACCTCAACTCGTGAAGAGAGCTATGGTG	855			
QY	1137	TGGAGAACAGACCTACTCTCAAGAGGAAATCACTCAAGGTTTGAAGAGACAGGGGACA	1196			
DB	856	TGGAGAACAGACCTACTCTCAAGAGGAAATCACTCAAGGTTTGAAGAGACAGGGGACA	915			
QY	1197	CGCTCTATGCCCTTATTCACACATTTTTCAGTGCAGAACCGCTCCCTCCCAAGGAT	1256			
DB	916	CGCTCTATGCCCTTATTCACACATTTTTCAGTGCAGAACCGCTCCCTCCCAAGGAT	975			
QY	1257	TCTCCATCCACGGGCCCCAGCTTGGCCGAGCCCTTTACAAAGACTATGAAGTAAAGAAAG	1316			
DB	976	TCTCCATCCACGGGCCCCAGCTTGGCCGAGCCCTTTACAAAGACTATGAAGTAAAGAAAG	1035			

QY	1317	AGGGCAGCTAACTCTGTCTCTGAAGAGTGGGACAAATGAGCCGGGGCGGAGATCTAGCGG	1376			
DB	1036	AGGGCAGCTAACTCTGTCTCTGAAGAGTGGGACAAATGAGCCGGGGCGGAGATCTAGCGG	1095			
QY	1377	GAGCTCAAAGGATGTGGCGAAATCTTGAGTCTTCTGAGAAAACCTGTACAAGACACTAC	1436			
DB	1096	GAGCTCAAAGGATGTGGCGAAATCTTGAGTCTTCTGAGAAAACCTGTACCTCCCGGGC	1155			
QY	1437	GG 1438				
DB	1156	GG 1157				
RESULT 6						
US-09-312-283C-249						
; Sequence 249, Application US/09312283C						
; Patent No. 6573095						
; GENERAL INFORMATION:						
; APPLICANT: Watson, James D.						
; APPLICANT: Strachan, Lorna						
; APPLICANT: Sleeman, Matthew						
; APPLICANT: Onrust, Rene						
; APPLICANT: Murison, James G.						
; APPLICANT: Kumble, Krishanand D.						
; TITLE OF INVENTION: Compositions Isolated from Skin Cells						
; FILE REFERENCE: 11000.1011c2						
; CURRENT APPLICATION NUMBER: US/09/312,283C						
; CURRENT FILING DATE: 1999-05-14						
; NUMBER OF SEQ ID NOS: 425						
; SOFTWARE: FastSeq for Windows Version 4.0						
; SEQ ID NO 249						
; LENGTH: 1212						
; TYPE: DNA						
; ORGANISM: Mouse						
US-09-312-283C-249						
Query Match 45.8%; Score 1126; DB 3; Length 1212;						
Best Local Similarity 99.1%; Pred. No. 2.8e-305;						
Matches 1132; Conservative 0; Mismatches 10; Indels 0; Gaps 0;						
QY	297	CCAAGTACTACAGACTTTGTGATAAGCTTGAAGCTTGGGGCATCGTCTAGAAAACGGTGG	356			
DB	16	CGAGGTACTACAGACTTTGTGATAAGCTTGAAGCTTGGGGCATCGTCTAGAAAACGGTGG	75			
QY	357	CCACAGCGGGGTGTGACCTCGGTGGCTTCATGCTCACTCTCCGATCCTCGTCTGCA	416			
DB	76	CCACAGCTGGGTGTGACCTCGGTGGCTTCATGCTCACTCTCCGATCCTCGTCTGCA	135			
QY	417	AGGTGAGGACTCCAAAGCGGAAAAATGCTGCTCACTCGCTACTCAGTTTCTTCTCTGGTG	476			
DB	136	AGGTGAGGACTCCAAAGCGGAAAAATGCTGCTCACTCGCTACTCAGTTTCTTCTCTGGTG	195			
QY	477	TGTTGGGCATCTTGGCCCTCAGCTTCGCTTTCATCATCGGACTGGACGGGACACAGGC	536			
DB	196	TGTTGGGCATCTTGGCCCTCAGCTTCGCTTTCATCATCGGACTGGACGGGACACAGGC	255			
QY	537	CCACAGCTTCTTCCCTTTGGGATCCTCTTTTCCAATCTGCTTCTCTCTCTGGCTC	596			
DB	256	CCACAGCTTCTTCCCTTTGGGATCCTCTTTTCCAATCTGCTTCTCTCTCTGGCTC	315			
QY	597	ATGCTGTCACTGACCAAGCTCGTCCGGGGAGGAGCCCTTTCCCTGTTGGTGATTC	656			
DB	316	ATGCTGTCACTGACCAAGCTCGTCCGGGGAGGAGCCCTTTCCCTGTTGGTGATTC	375			
QY	657	TGGGTCTGGCCGTGAGCTTACAGGTAGTCCAGGATGTTATCGCTATTGAATATATTGTC	716			
DB	376	TGGGTCTGGCCGTGAGCTTACAGGTAGTCCAGGATGTTATCGCTATTGAATATATTGTC	435			
QY	717	TGACCATGAATAGGACCAACGTCAATGTCTTTTCTGAGCTTTCCGCTCTCGTGGCAATG	776			
DB	436	TGACCATGAATAGGACCAACGTCAATGTCTTTTCTGAGCTTTCCGCTCTCGTGGCAATG	495			



777 AAGACTTTGTCCTCCTGCTCAGCTACGTCCTCTTCTTGATGGCGCTGACCTCTCTCATGT 836  
Db |||||  
496 AAGACTTTGTCCTCCTGCTCAGCTACGTCCTCTTCTTGATGGCGCTGACCTCTCTCATGT 555  
Qy |||||  
837 CTTCTTACCTTCTGTTGTTCTTCAACGGGCTGGAAGAGACATGGGGCCACATCTACC 896  
Db |||||  
556 CTTCTTACCTTCTGTTGTTCTTCAACGGGCTGGAAGAGACATGGGGCCACATCTACC 615  
Qy |||||  
897 TCAGGATGCTCCTTCCATGTCCTTCACTGCTGCTGGGTGGCTGGATCAGCTGCTTCTG 956  
Db |||||  
616 TCAGGATGCTCCTTCCATGTCCTTCACTGCTGCTGGGTGGCTGGATCAGCTGCTTCTG 675  
Qy |||||  
957 ACTTTGACCCAGGTGGGATGACACCATCTCAGCTCCGCTTGGCTGGCTGCAATGGCTGGG 1016  
Db |||||  
676 ACTTTGACCCAGGTGGGATGACACCATCTCAGCTCCGCTTGGCTGGCTGCAATGGCTGGG 735  
Qy |||||  
1017 TGTCTCTGTTGGCTTATGTTAGTCCCGAGTTTGGCTGCTCACAAGCAACGAAACCCCA 1076  
Db |||||  
736 TGTCTCTGTTGGCTTATGTTAGTCCCGAGTTTGGCTGCTCACAAGCAACGAAACCCCA 795  
Qy |||||  
1077 TGGATTATCTGTTGAGGATGCTTCTGTAACCTCAACTCGTGAAGAGAGCTATGGTG 1136  
Db |||||  
796 TGGATTATCTGTTGAGGATGCTTCTGTAACCTCAACTCGTGAAGAGAGCTATGGTG 855  
Qy |||||  
1137 TGGAGAACAGAGCTTACTCTCAAGAGGAAATCACTCAAGGTTTGAAGAGACAGGGGACA 1196  
Db |||||  
856 TGGAGAACAGAGCTTACTCTCAAGAGGAAATCACTCAAGGTTTGAAGAGACAGGGGACA 915  
Qy |||||  
1197 CGCTTATGCCCTTATTCACACATTTTCAGCTGCGAGAACCCAGCTTCCCAAAAGGAAT 1256  
Db |||||  
916 CGCTTATGCCCTTATTCACACATTTTCAGCTGCGAGAACCCAGCTTCCCAAAAGGAAT 975  
Qy |||||  
1257 TCTCCATCCACGGGCCAGCTTGGCCGAGCCCTTACAAGACTATGAAGTAAAGAAAG 1316  
Db |||||  
976 TCTCCATCCACGGGCCAGCTTGGCCGAGCCCTTACAAGACTATGAAGTAAAGAAAG 1035  
Qy |||||  
1317 AGGGCAGCTAACTCTGCTGGAAGAGTGGGACAAATGCAGCCGGCGGAGATCTAGCGG 1376  
Db |||||  
1036 AGGGCAGCTAACTCTGCTGGAAGAGTGGGACAAATGCAGCCGGCGGAGATCTAGCGG 1095  
Qy |||||  
1377 GAGCTCAAAAGGGATGTGGGCGAAATCTTGAGTCTTCTGAGAAACTGTACAAGACACTAC 1436  
Db |||||  
1096 GAGCTCAAAAGGGATGTGGGCGAAATCTTGAGTCTTCTGAGAAACTGTACTCCCGGGC 1155  
Qy |||||  
1437 GG 1438  
Db |||||  
1156 GG 1157

RESULT 7  
US-09-495-050A-223  
; Sequence 223, Application US/09495050A  
; Patent No. 6492505  
; GENERAL INFORMATION:  
; APPLICANT: Roopa, Reddy  
; APPLICANT: Guegler, Karl, J.  
; APPLICANT: Au-Young, Janice  
; TITLE OF INVENTION: COMPOSITION FOR DETECTION OF GENES ENCODING MEMBRANE-ASSOCIATED P  
; FILE REFERENCE: PA-0013 US  
; CURRENT FILING DATE: 2000-01-31  
; PRIOR APPLICATION NUMBER: 60/118,318  
; PRIOR FILING DATE: February 1, 1999  
; NUMBER OF SEQ ID NOS: 305  
; SOFTWARE: PERL Program  
; SEQ ID NO 223  
; LENGTH: 1228  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: Incyte ID No. 6492505 2681738CT1  
US-09-495-050A-223

Query Match 42.0%; Score 1031.6; DB 3; Length 1228;  
Best Local Similarity 98.6%; Pred. No. 8.3e-279;  
Matches 1094; Conservative 0; Mismatches 9; Indels 7; Gaps 5;  
Qy 732 CCAACGTCAATGCTTTTCTGAGCTTTCCGCTCCTCGTCGCAATGAAGACTTTTGTCTCTCC 791  
Db |||||  
1 CCAACGTCAATGCTTTTCTGAGCTTTCCGCTCCTCGTCGCAATGAAGACTTTTGTCTCTCC 60  
Qy 792 TGTCTACCTACGTCCTTCTTCTGATGGCGCTGACCTTCTCTCATGTCCTCTTCACTTCT 851  
Db |||||  
61 TGTCTACCTACGTCCTTCTTCTGATGGCGCTGACCTTCTCTCATGTCCTCTTCACTTCT 120  
Qy 852 GTGGTTCCTTCAACGGCTGGAAGAGACATGGGGCCACATCTACCTACGATGCTCTCT 911  
Db |||||  
121 GTGGTTCCTTCAACGGCTGGAAGAGACATGGGGCCACATCTACCTACGATGCTCTCT 180  
Qy 912 CCATTGCCATCTGGGTGGCTGGATCACCTCGTCTCATGCTTCTGACTTTTGACCGCAGGT 971  
Db |||||  
181 CCATTGCCATCTGGGTGGCTGGATCACCTCGTCTCATGCTTCTGACTTTTGACCGCAGGT 240  
Qy 972 G-GGATGACACCATCTCAGCTCCGCTTGGCTGGCAATGGCTGGGTGGTCTTCTGCTTGGCT 1030  
Db |||||  
241 GAGGATGACACCATCTCAGCTCCGCTTGGCTGGCAATGGCTGGGTGGTCTTCTGCTTGGCT 300  
Qy 1031 TATGTTAGTCCCGAGTTTGGCTGCTCACAAGCAACGAAACCCCATGGATTAATCTCTGTT 1090  
Db |||||  
301 TATGTTAGTCCCGAGTTTGGCTGCTCACAAGCAACGAAACCCCATGGATTAATCTCTGTT 360  
Qy 1091 GAGGATGCTTTTCTGTAACCTCAACTCGTGAAGAGAGCTATGGTGTGGAGAACAGAGCC 1150  
Db |||||  
361 GAGGATGCTTTTCTGTAACCTCAACTCGTGAAGAGAGCTATGGTGTGGAGAACAGAGCC 420  
Qy 1151 TACTCTCAAGAGGAAATCACTCAAGGTTTGAAGAGACAGGGGACACGCTCTATGCCCCC 1210  
Db |||||  
421 TACTCTCAAGAGGAAATCACTCAAGGTTTGAAGAGACAGGGGACACGCTCTATGCCCCC 480  
Qy 1211 TATTTCCACACATTTTCAAGTGCAGAACCCAGCTTCCCAAGAAATTTCTCCATCCCAAGG 1270  
Db |||||  
481 TATTTCCACACATTTTCAAGTGCAGAACCCAGCTTCCCAAGAAATTTCTCCATCCCAAGG 540  
Qy 1271 GCCACGCTTGGCGAGCCCTTACAAGACTATGAAGTAAAGAAAGAGGCGAGCTAACTC 1330  
Db |||||  
541 GCCACGCTTGGCGAGCCCTTACAAGACTATGAAGTAAAGAAAGAGGCGAGCTAACTC 600  
Qy 1331 TGTCTGAAAGAGTGGGACAAATGCAGCGGGCGGAGATCTAGCGGAGCTCAAGGGAT 1390  
Db |||||  
601 TGTCTGAAAGAGTGGGACAAATGCAGCGGGCGGAGATCTAGCGGAGCTCAAGGGAT 660  
Qy 1391 GTGGGCGAAATCTTGAGTCTTCTGAGAAACTGTACAAGACACTACGGGAAACAGTTTGGCC 1450  
Db |||||  
661 GTGGGCGAAATCTTGAGTCTTCTGAGAAACTGTACAAGACACTACGGGAAACAGTTTGGCC 720  
Qy 1451 TCCCTCCAGCTTCAACCAATTTCTTCAATGCTGGGGCTGATGTGGGCTAGTAAGACTC 1510  
Db |||||  
721 TCCCTCCAGCTTCAACCAATTTCTTCAATGCTGGGGCTGATGTGGGCTAGTAAGACTC 780  
Qy 1511 CAGTTCTTAGAGCGCTGAGTA - TTTTCTTTTCTGCTCATCTCTTTGGATACCTTCT 1568  
Db |||||  
781 CAGTTCTTAGAGCGCTGAGTA - TTTTCTTTTCTGCTCATCTCTTTGGATACCTTCT 840  
Qy 1569 TTTAAGTGGGAGTCTCAGGCAACTCAAGTTTACACCTTACTCTTTTGTGTTTTTGTG 1628  
Db |||||  
841 TTTAAGTGGGAGTCTCAGGCAACTCAAGTTTACACCTTACTCTTTTGTGTTTTTGTG 900  
Qy 1629 AAAACAGGATCTTGTCTGTGACCCAGGCTTGGTGGAGTGGTGGATCAGAGCCAGTGC 1688  
Db |||||  
901 AAAACAGGATCTTGTCTGTGACCCAGGCTTGGTGGAGTGGTGGATCAGAGCCAGTGC 960  
Qy 1689 AGCTTCGACCACTGTGCTCAAGCAATCTCTCCATCTCCCAAGTGTGGGATG 1748  
Db |||||  
961 AGCTTCGACCACTGTGCTCAAGCAATCTCTCCATCTCTCCCAAGTGTGGGATG 1020



Db 481 AAAGATCACCTAAATTTACTCATCTCTAGTGTGCTCCATCAGCTGAGGCTCAGCAGCT 540  
QY 2090 CCCAGACCAATTCAGGTACCCCTCTCTCTTCTGCACTGTCCCAAACTTGTGTCA 2149  
Db 541 CCCAGACCAATTCAGGTACCCCTCTCTCTTCTGCACTGTCCCAAACTTGTGTCA 600  
QY 2150 A 2150  
Db 601 A 601

RESULT 10  
US-09-949-016-126957  
; Sequence 126957, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 126957  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-126957

Query Match 24.5%; Score 600.6; DB 3; Length 601;  
Best Local Similarity 99.8%; Pred. No. 3.9e-158;  
Matches 600; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1258 CTCATCCACGGGCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAGAAAGA 1317  
Db 1 CTCATCCACGGGCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAGAAAGA 60

QY 1318 GGGCAGCTAACTCTGTCTGAGAGTGGGCAAAATCAGCGCGGCGGAGATCAGCGGG 1377  
Db 61 GGGCAGCTAACTCTGTCTGAGAGTGGGCAAAATCAGCGCGGCGGAGATCAGCGGG 120

QY 1378 AGCTCAAGGGATGTGGGCGAAATCTTGAGTCTTCTGAGAAAATCTGACAAAGACTACG 1437  
Db 121 AGCTCAAGGGATGTGGGCGAAATCTTGAGTCTTCTGAGAAAATCTGACAAAGACTACG 180

QY 1438 GGAACAGTTTGCTCCCTCCAGCCTCAACCAAAATCTTCCATGCTGGGGCTGATGTGG 1497  
Db 181 GGAACAGTTTGCTCCCTCCAGCCTCAACCAAAATCTTCCATGCTGGGGCTGATGTGG 240

QY 1498 GCTAGTAAGACTCCAGTCTTAGAGCGCTGTAGTATTTTTTTTTTGTCTCATCCTT 1557  
Db 241 GCTAGTAAGACTCCAGTCTTAGAGCGCTGTAGTATTTTTTTTTTGTCTCATCCTT 300

QY 1558 TGGATACTTTCTTTTAAAGTGGAGTCTCAGCAACTCAAGTTTAGACCCCTTACTCTTTTGG 1617  
Db 301 WGGATACTTTCTTTTAAAGTGGAGTCTCAGGCAACTCAAGTTTAGACCCCTTACTCTTTTGG 360

QY 1618 TTTGTTTTTGAACAGGATCTTGTCTGTGACCCAGGCTTGAAGTGGTGGCATCA 1677  
Db 361 TTTGTTTTTGAACAGGATCTTGTCTGTGACCCAGGCTTGAAGTGGTGGCATCA 420

QY 1678 CAGCCAGTGCAGCTCGACCACCTGTCTCAAGCAATCTCCCATCTCCATCTCCAAA 1737  
Db 421 CAGCCAGTGCAGCTCGACCACCTGTCTCAAGCAATCTCCCATCTCCATCTCCAAA 480

QY 1738 GTGCTGGGATGACAGGCGTGAGCCACAGCTCCCGAGCTAGGCCCTTAAATCTTGTGTAT 1797

Db 481 GTGCTGGGATGACAGGCTGAGCCACAGCTCCCGAGCTTAATCTTGTGTAT 540  
QY 1798 TTTCCATGAGCTAAAGGTCTGTGTCATCTAGCTCAGCTGGCTCACACAGCTCTAGGGC 1857  
Db 541 TTTCCATGAGCTAAAGGTCTGTGTCATCTAGCTCAGCTGGCTCACACAGCTCTAGGGC 600  
QY 1858 C 1858  
Db 601 C 601

RESULT 11  
US-09-949-016-126958  
; Sequence 126958, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 126958  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-126958

Query Match 24.5%; Score 600.6; DB 3; Length 601;  
Best Local Similarity 99.8%; Pred. No. 3.9e-158;  
Matches 600; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1550 TCATCCTTTGGTACTTCTTTTAAAGTGGGAGTCTCAGGCAACTCAAGTTTAGACCCCTTAC 1609  
Db 1 TCATCCTTTGGTACTTCTTTTAAAGTGGGAGTCTCAGGCAACTCAAGTTTAGACCCCTTAC 60

QY 1610 TCTTTTGTGTGTGTGTTTGAACAGGATCTTGTCTGTACCCAGGCTTGAAGTGCAGTGG 1669  
Db 61 TCTTTTGTGTGTGTGTTTGAACAGGATCTTGTCTGTACCCAGGCTTGAAGTGCAGTGG 120

QY 1670 TGGGATCAGACCCAGTGCAGCTCCGACCACTGTGCTCAAGCAATCTCCCATCTCCAT 1729  
Db 121 TGGGATCAGACCCAGTGCAGCTCCGACCACTGTGCTCAAGCAATCTCCCATCTCCAT 180

QY 1730 CTCCTCAAGTGTGGGATGACAGGCTGAGCCACAGCTCCAGCCTAGGCCCTTAATCTT 1789  
Db 181 CTCCTCAAGTGTGGGATGACAGGCTGAGCCACAGCTCCAGCCTAGGCCCTTAATCTT 240

QY 1790 GCTGTTATTTTCATGGACTTAAAGGTCTGGTCACTGAGCTACGCTGGCTCAGACAGT 1849  
Db 241 GCTGTTATTTTCATGGACTTAAAGGTCTGGTCACTGAGCTACGCTGGCTCAGACAGT 300

QY 1850 CTAGGGGCTGTCTCTTAACCTCACAGTGGGTTTGTGAGGCTCTGTGGCCAGAGCAGA 1909  
Db 301 YTAGGGGCTGTCTCTCTTAACCTCACAGTGGGTTTGTGAGGCTCTGTGGCCAGAGCAGA 360

QY 1910 CTTGCATATCTGAGCAAAAATAGCAAAAGCTCTCTCAGCCCACTGGCCTGAATCTACAC 1969  
Db 361 CTTGCATATCTGAGCAAAAATAGCAAAAGCTCTCTCAGCCCACTGGCCTGAATCTACAC 420

QY 1970 TGAAGCCAACTTGTGTGGCACCCCGCTCCCAACCCCTTCTGCTGGGTAGGAGGCT 2029  
Db 421 TGAAGCCAACTTGTGTGGCACCCCGCTCCCAACCCCTTCTGCTGGGTAGGAGGCT 480



; SEQ ID NO 4  
; LENGTH: 311  
; TYPE: DNA  
; ORGANISM: Human  
US-09-188-930-4

Query Match 11.9%; Score 292.2; DB 3; Length 311;  
Best Local Similarity 99.0%; Pred. No. 8.1e-72;  
Matches 294; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

195 TGTGGGCATCTTTGGCCTCACCTTCGCTTCATCGGACTGGAGCGGAGCACAGGC 254  
537 CCACACGCTTCTCTCTTTGGGATCCTCTTTTCCATCTGCTTCTCTCTGCTGCTGG 593  
255 CCACACGCTTCTCTCTTTGGGATCCTCTTTTCCATCTGCTTCTCTCTGCTGCTGG 311

Search completed: June 10, 2006, 15:26:30  
Job time : 433 secs

297 CCAAGTACTACAGACTTTGTGATAAGGCTGAAGCTTTGGGGCATCGTCTAGAAACGGTGG 356  
15 CGAGGTACTACAGACTTTGTGATAAGGCTGAAGCTTTGGGGCATCGTCTAGAAACGGTGG 74  
357 CCACAGCGGGGTGTGACCTCGGTGGCTTCAATGCTCACTCTCCGATCCTCGTCTGCA 416  
75 CCACAGTGGGGTGTGACCTCGGTGGCTTCAATGCTCACTCTCCGATCCTCGTCTGCA 134  
417 AGGTGAGGACTCCAAACAGCGGAAAAATGCTGCTACTCAGTTTCTCTCTCTCTGGGTG 476  
135 AGGTGAGGACTCCAAACAGCGGAAAAATGCTGCTACTCAGTTTCTCTCTCTGGGTG 194  
477 TGTGGGCATCTTTGGCCTCACCTTCGCTTCAATCGGACTGGAGCGGAGCACAGGC 536  
195 TGTGGGCATCTTTGGCCTCACCTTCGCTTCAATCGGACTGGAGCGGAGCACAGGC 254  
537 CCACACGCTTCTCTCTTTGGGATCCTCTTTTCCATCTGCTTCTCTGCTGCTGG 593  
255 CCACACGCTTCTCTCTTTGGGATCCTCTTTTCCATCTGCTTCTCTGCTGCTGG 311

RESULT 15  
US-09-312-283C-4  
; Sequence 4, Application US/09312283C  
; Patent No. 6573095  
; GENERAL INFORMATION:  
; APPLICANT: Watson, James D.  
; APPLICANT: Strachan, Lorna  
; APPLICANT: Sleeman, Matthew  
; APPLICANT: Onrust, Rene  
; APPLICANT: Murison, James G.  
; APPLICANT: Kumble, Krishanand D.  
; TITLE OF INVENTION: Compositions Isolated from Skin Cells  
; FILE REFERENCE: 11000.1011c2  
; CURRENT APPLICATION NUMBER: US/09/312,283C  
; CURRENT FILING DATE: 1999-05-14  
; NUMBER OF SEQ ID NOS: 425  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 311  
; TYPE: DNA  
; ORGANISM: Mouse  
US-09-312-283C-4

Query Match 11.9%; Score 292.2; DB 3; Length 311;  
Best Local Similarity 99.0%; Pred. No. 8.1e-72;  
Matches 294; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

297 CCAAGTACTACAGACTTTGTGATAAGGCTGAAGCTTTGGGGCATCGTCTAGAAACGGTGG 356  
15 CGAGGTACTACAGACTTTGTGATAAGGCTGAAGCTTTGGGGCATCGTCTAGAAACGGTGG 74  
357 CCACAGCGGGGTGTGACCTCGGTGGCTTCAATGCTCACTCTCCGATCCTCGTCTGCA 416  
75 CCACAGTGGGGTGTGACCTCGGTGGCTTCAATGCTCACTCTCCGATCCTCGTCTGCA 134  
417 AGGTGAGGACTCCAAACAGCGGAAAAATGCTGCTACTCAGTTTCTCTCTCTGGGTG 476  
135 AGGTGAGGACTCCAAACAGCGGAAAAATGCTGCTACTCAGTTTCTCTCTCTGGGTG 194  
477 TGTGGGCATCTTTGGCCTCACCTTCGCTTCAATCGGACTGGAGCGGAGCACAGGC 536  
255 CCACACGCTTCTCTCTTTGGGATCCTCTTTTCCATCTGCTTCTCTGCTGCTGG 311

THIS PAGE BLANK (USPTO)

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2006, 15:08:58 ; Search time 3003 Seconds  
(without alignments)  
10049.428 Million cell updates/sec

Title: US-10-600-816-2

Perfect score: 2456

Sequence: 1 ataacagcatgaagtcgcgt.....ggcagcaaaaaaaaaa 2456

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 18892170 seqs, 6143817638 residues

Total number of hits satisfying chosen parameters: 37784340

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA Main:\*

- 1: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*
- 2: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*
- 3: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*
- 4: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*
- 5: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*
- 6: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*
- 7: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*
- 8: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10C\_PUBCOMB.seq.\*
- 9: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10D\_PUBCOMB.seq.\*
- 10: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10E\_PUBCOMB.seq.\*
- 11: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10F\_PUBCOMB.seq.\*
- 12: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10G\_PUBCOMB.seq.\*
- 13: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US11A\_PUBCOMB.seq.\*
- 14: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US11B\_PUBCOMB.seq.\*
- 15: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US11C\_PUBCOMB.seq.\*
- 16: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US11D\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2456	100.0	2456	6	US-10-225-567A-453
2	2456	100.0	2456	7	US-10-269-909-63
3	2456	100.0	2456	7	US-10-269-909-64
4	2456	100.0	2456	7	US-10-295-027-619
5	2456	100.0	2456	8	US-10-600-816-2
6	2456	100.0	2456	9	US-10-775-920-10
7	2456	100.0	2456	10	US-10-936-626-40
8	2456	100.0	2456	10	US-10-938-061-40
9	2456	100.0	2456	16	US-11-169-041-32
10	2448	99.5	4239	6	US-10-600-816-18
11	2439	99.3	4239	6	US-10-198-846-10424
12	2439	99.3	2302	7	US-10-775-920-11
13	2302	93.7	2302	7	US-10-224-289-3
14	2302	93.7	2302	8	US-10-240-425-405
15	2302	93.7	2302	9	US-10-775-920-9
16	2302	93.7	2302	10	US-10-510-507-2
17	2286	93.1	2316	6	US-10-176-847-59

18	2286	93.1	2316	15	US-11-080-991-59	Sequence 59, Appl
19	2274.8	92.6	2297	9	US-10-775-920-13	Sequence 13, Appl
20	2268.2	92.4	2305	9	US-10-775-920-12	Sequence 12, Appl
21	2260.4	92.0	2593	7	US-10-264-049-834	Sequence 834, App
22	1601.8	65.2	1619	7	US-10-224-289-5	Sequence 5, Appl1
23	1601.8	65.2	1619	9	US-10-935-190-43	Sequence 43, Appl1
24	1441.4	58.7	1460	10	US-10-936-626-64	Sequence 64, Appl
25	1441.4	58.7	1460	10	US-10-938-061-64	Sequence 64, Appl
26	1400	57.0	1400	13	US-11-060-756-2418	Sequence 2418, Ap
27	1400	57.0	1400	13	US-11-060-756-6690	Sequence 6690, Ap
28	1126	45.8	1212	3	US-09-866-050A-249	Sequence 249, App
29	1126	45.8	1212	6	US-10-152-661-210	Sequence 210, App
30	1082.6	44.1	1114	3	US-09-978-360A-210	Sequence 210, App
31	1082.6	44.1	1114	5	US-09-978-360A-210	Sequence 210, App
32	1071	43.6	1071	9	US-10-712-615-134	Sequence 134, App
33	1067.8	43.5	1788	10	US-10-505-486-196	Sequence 196, App
34	1031.6	42.0	1228	7	US-10-313-542-223	Sequence 223, App
C 35	932.6	38.0	1000	3	US-09-864-761-19238	Sequence 19238, A
C 36	787.6	32.1	948	3	US-09-864-761-2510	Sequence 2510, Ap
C 37	702.4	28.6	774	6	US-10-106-698-2079	Sequence 2079, Ap
C 38	661.2	26.9	1934	7	US-10-224-289-1	Sequence 1, Appli
39	602	24.5	715	3	US-09-969-034-1270	Sequence 1270, Ap
C 40	540.4	22.0	620	3	US-09-969-034-2222	Sequence 2222, Ap
C 41	538.4	21.9	552	6	US-10-066-543-1811	Sequence 1811, Ap
42	497	20.2	497	6	US-10-066-543-1937	Sequence 1937, Ap
C 43	496.6	20.2	562	3	US-09-969-034-1808	Sequence 1808, Ap
C 44	477.2	19.4	519	6	US-10-052-283-396	Sequence 396, App
C 45	467	19.0	468	3	US-09-998-598-2451	Sequence 2451, Ap

ALIGNMENTS

RESULT 1

US-10-225-567A-453  
; Sequence 453, Application US/10225567A  
; Publication No. US20030113798A1  
; GENERAL INFORMATION:  
; APPLICANT: Lifespan Biosciences  
; APPLICANT: Brown, Joseph P.  
; APPLICANT: Burner, Glenna C.  
; APPLICANT: Roush, Christine L.  
; TITLE OF INVENTION: ANTIGENIC PEPTIDES AND ANTIBODIES FOR G PROTEIN-COUPLED RECEPTORS  
; FILE REFERENCE: 1920-4-4  
; CURRENT APPLICATION NUMBER: US/10/225,567A  
; PRIOR FILING DATE: 2001-12-19  
; PRIOR APPLICATION NUMBER: 60/257,144  
; NUMBER OF SEQ ID NOS: 2292  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 453  
; LENGTH: 2456  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-225-567A-453

Query Match	100.0%	Score 2456;	DB 6;	Length 2456;
Best Local Similarity	100.0%;	Pred. No. 0.		
Matches 2456;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	ATAACAGCATGAAGTCGCGTGAACCTGGAATAGGCGGTCTCTCCCTCGACCCCTCCCC	60	
Db	1	ATAACAGCATGAAGTCGCGTGAACCTGGAATAGGCGGTCTCTCCCTCGACCCCTCCCC	60	
QY	61	TCCTTGTCCTCTGCTACCCCTCGCTTCCTCCCTCCGCGAGGGCGGCGCTTTATA	120	
Db	61	TCCTTGTCCTCTGCTACCCCTCGCTTCCTCCCTCCGCGAGGGCGGCGCTTTATA	120	
QY	121	ACAACTGCTCAGATGCGAGGGGGATAGCTGTTCCTCCCTCCGCGAGGGCGGCGCTTTATA	180	
Db	121	ACAACTGCTCAGATGCGAGGGGGATAGCTGTTCCTCCCTCCGCGAGGGCGGCGCTTTATA	180	
QY	181	CTCGCTGCTGCCCTCTTTCGCGGGGAAGCAGCACCAGTTTCACGGCCAAACGCTTGGC	240	

181 CTCGGCTGTCCTCTTCGCGCGGGAAGAGACCAAGATTACGGCCAAACGCTTGGC 240  
241 ACTAGGGTCCAGATGGCTACAACAGTCCCTGATGGTTGCCGCAATGGGCTGAATCCAA 300  
241 ACTAGGGTCCAGATGGCTACAACAGTCCCTGATGGTTGCCGCAATGGGCTGAATCCAA 300  
301 GPTACTACAGACTTTGTGATAAGGCTGAAGCTTTGGGCAATCGTCTTAGAAACCGGTGGCCAC 360  
301 GPTACTACAGACTTTGTGATAAGGCTGAAGCTTTGGGCAATCGTCTTAGAAACCGGTGGCCAC 360  
361 AGCCGGGTTGTACCTCGGTGGCCTTATGCTCACTCTCCGATCCTGCTGCAAGT 420  
361 AGCCGGGTTGTACCTCGGTGGCCTTATGCTCACTCTCCGATCCTGCTGCAAGT 420  
421 GCAGAGACTCCAAACAGGCGAAAAATGCTGCTACTCAGTTTCTCTTCTCTCGGTGTGTT 480  
421 GCAGAGACTCCAAACAGGCGAAAAATGCTGCTACTCAGTTTCTCTTCTCTCGGTGTGTT 480  
481 GGGCATCTTTGGCCTCACTTGGCCTTCAATCGGAAGTGAACGGGAGACAGGGGCCAC 540  
481 GGGCATCTTTGGCCTCACTTGGCCTTCAATCGGAAGTGAACGGGAGACAGGGGCCAC 540  
541 AGCTTCTCTCTTTGGGATCCTCTTTTCCATCTGCTTCTCTGCTGCTGGCTCATGC 600  
541 AGCTTCTCTCTTTGGGATCCTCTTTTCCATCTGCTTCTCTGCTGCTGGCTCATGC 600  
601 TGTGAGTCTGACCAAGCTCGTCCGGGGAGGAGCCCTTCCCTGTTGGTGAATCTGGG 660  
601 TGTGAGTCTGACCAAGCTCGTCCGGGGAGGAGCCCTTCCCTGTTGGTGAATCTGGG 660  
661 TCTGGCCGTGGGCTCAGCCTAGTCAGGATGTTATCGTATATGAATATATTTGCTGAC 720  
661 TCTGGCCGTGGGCTCAGCCTAGTCAGGATGTTATCGTATATGAATATATTTGCTGAC 720  
721 CATGATAGGACCAAGCTCAATGCTCTTTTCTCAGCTTTCCGCTCCGTCGCAATGAAGA 780  
721 CATGATAGGACCAAGCTCAATGCTCTTTTCTCAGCTTTCCGCTCCGTCGCAATGAAGA 780  
781 CTTTGTCTCTCTGCTCACCCTAGCTCTCTTTGATGGGCTGACCTTCTCATGTCCTC 840  
781 CTTTGTCTCTCTGCTCACCCTAGCTCTCTTTGATGGGCTGACCTTCTCATGTCCTC 840  
841 CTTACCTTCTGTGTTTCCTTCAOGGGCTGGAAGAGACATGGGGCCACATCTACTCAC 900  
841 CTTACCTTCTGTGTTTCCTTCAOGGGCTGGAAGAGACATGGGGCCACATCTACTCAC 900  
901 GATGCTCTCTCCATTGGCCATCTGGGTGGCTGGATCACCTGCTCATGCTTCTCGACTT 960  
901 GATGCTCTCTCCATTGGCCATCTGGGTGGCTGGATCACCTGCTCATGCTTCTCGACTT 960  
961 TGACCGCAGTGGGATGACACCATCTCTCAGCTCCGCTTGGCTGGCAATGGGTGGTGT 1020  
961 TGACCGCAGTGGGATGACACCATCTCTCAGCTCCGCTTGGCTGGCAATGGGTGGTGT 1020  
1021 CTTGTTGGCTTATGTTAGTCCGAGTTTGGCTGTCTCAAAAGCAACGAAACCCCATGGA 1080  
1021 CTTGTTGGCTTATGTTAGTCCGAGTTTGGCTGTCTCAAAAGCAACGAAACCCCATGGA 1080  
1081 TTATCTGTTGAGGATGCTTTCTGTAAACCTCAACTCGTGAAGAGAGCTATGGTGTGA 1140  
1081 TTATCTGTTGAGGATGCTTTCTGTAAACCTCAACTCGTGAAGAGAGCTATGGTGTGA 1140  
1141 GAACGAGCCTACTCTCAAGAGAAATCACTCAAGGTTTTGAAGAGACAGGGGACACGCT 1200  
1141 GAACGAGCCTACTCTCAAGAGAAATCACTCAAGGTTTTGAAGAGACAGGGGACACGCT 1200  
1201 CTATGCCCCCTATTCACACATTTTCAGCTGAGAAACAGCCTCCCAAAAGAAATTTCTC 1260  
1201 CTATGCCCCCTATTCACACATTTTCAGCTGAGAAACAGCCTCCCAAAAGAAATTTCTC 1260  
1261 CATCCACGGGCCAGGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAGAAAGAGGG 1320

1261 CATCCACGGGCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAGAAAGAGGG 1320  
1321 CAGCTAACTCTGTCTCTGAAGAGTGGGACAAAATGCAGCCGGGCGGAGACTTCTAGCGGAGC 1380  
1321 CAGCTAACTCTGTCTCTGAAGAGTGGGACAAAATGCAGCCGGGCGGAGACTTCTAGCGGAGC 1380  
1381 TCAAAAGGATGTGGCGGAAAATCTTGAGTCTTCTGAGAAAATGTACAAGACACTACGGGA 1440  
1381 TCAAAAGGATGTGGCGGAAAATCTTGAGTCTTCTGAGAAAATGTACAAGACACTACGGGA 1440  
1441 ACAGTTTGCCTCCCTCCAGCCCTCAACACAATTTCTTCCATCTGGGGCTGATGTGGGT 1500  
1441 ACAGTTTGCCTCCCTCCAGCCCTCAACACAATTTCTTCCATCTGGGGCTGATGTGGGT 1500  
1501 AGTAAGACTCCAGTTCTTTAGAGGCGCTGTAGTATTTTTTTTTTTTTTTTGTCTCATCTTGG 1560  
1501 AGTAAGACTCCAGTTCTTTAGAGGCGCTGTAGTATTTTTTTTTTTTTTTTGTCTCATCTTGG 1560  
1561 ATACTTCTTTTAAGTGGGAGTCTCAGGCAATCAAGATTTAGACCCCTTACTCTTTTGTGTT 1620  
1561 ATACTTCTTTTAAGTGGGAGTCTCAGGCAATCAAGATTTAGACCCCTTACTCTTTTGTGTT 1620  
1621 GTTTTTGAAAACAGGATCTTGTCTGTCAACCAGGCTTGAGTGAGTGCGATCACAG 1680  
1621 GTTTTTGAAAACAGGATCTTGTCTGTCAACCAGGCTTGAGTGAGTGCGATCACAG 1680  
1681 CCCAGTGCAGCTCGAACCACTGTCTCAAGCAATCTCCCATCTCCAAAGTG 1740  
1681 CCCAGTGCAGCTCGAACCACTGTCTCAAGCAATCTCCCATCTCCAAAGTG 1740  
1741 CTGGGATCAGAGGCTGAGCCACAGCTCCAGCTCAGGCTTAGGCCCTTAATCTGTGTTATTTT 1800  
1741 CTGGGATCAGAGGCTGAGCCACAGCTCCAGCTCAGGCTTAGGCCCTTAATCTGTGTTATTTT 1800  
1801 CCATGCACTAAAGTCTGCTCATCTGAGCTCAGCTGCTCACACAGCTCTTAGGGCCCTG 1860  
1801 CCATGCACTAAAGTCTGCTCATCTGAGCTCAGCTGCTCACACAGCTCTTAGGGCCCTG 1860  
1861 CTCCTCTAACTCAGTGGGTTTGTGAGGCTCTGTGGCCAGAGCAGACCTGCAATCT 1920  
1861 CTCCTCTAACTCAGTGGGTTTGTGAGGCTCTGTGGCCAGAGCAGACCTGCAATCT 1920  
1921 GAGCAAAAATAGCAAAAGCCTCTCTCAGCCCACTGGCCTGAATCTACACTGGAAGCCAAAC 1980  
1921 GAGCAAAAATAGCAAAAGCCTCTCTCAGCCCACTGGCCTGAATCTACACTGGAAGCCAAAC 1980  
1981 TTGCTGGCACCCCGCTCCCAACCCCTTCTTGCTGGGTAGGAGAGGCTTAAAGATCACCC 2040  
1981 TTGCTGGCACCCCGCTCCCAACCCCTTCTTGCTGGGTAGGAGAGGCTTAAAGATCACCC 2040  
2041 TAAATTTACTCATCTCTTAGTGTGCTTCAATTTGGGCTCAGCAGTCTCCCAAGACCA 2100  
2041 TAAATTTACTCATCTCTTAGTGTGCTTCAATTTGGGCTCAGCAGTCTCCCAAGACCA 2100  
2101 ATTACAGGTCACCCCTCTCTTCTTTCGACTGTCGCCCAACTTGTCTCAATTTCCGAGATC 2160  
2101 ATTACAGGTCACCCCTCTCTTCTTTCGACTGTCGCCCAACTTGTCTCAATTTCCGAGATC 2160  
2161 TAAATCTCCCTTACGCTCTGCGAGGAAATTTCTTTCAGACCTCACTAGCACAAGCCCGGTTG 2220  
2161 TAAATCTCCCTTACGCTCTGCGAGGAAATTTCTTTCAGACCTCACTAGCACAAGCCCGGTTG 2220  
2221 CTCCTTGTACAGAGAAATTTGTAGATCAATTTCAATTTCTGCGGCTGATCTTCT 2280  
2221 CTCCTTGTACAGAGAAATTTGTAGATCAATTTCAATTTCTGCGGCTGATCTTCT 2280  
2281 CTCATCTTGCAACCCCAACCTCTGTAATAGATTTTACCGCATTTACGGCTGCAATCTGTA 2340  
2281 CTCATCTTGCAACCCCAACCTCTGTAATAGATTTTACCGCATTTACGGCTGCAATCTGTA 2340  
2341 GTGGGCATGGTCTCCTTAATGGAGGAGTGTTCATTTGTATAATTAAGTTATTTCACCTGAGTAT 2400  
2341 GTGGGCATGGTCTCCTTAATGGAGGAGTGTTCATTTGTATAATTAAGTTATTTCACCTGAGTAT 2400







QY 961 TGACCGAGGTGGATGACACCATCTCAGCTCGGCTTGGCTGCCAATGGCTGGGTGTT 1020  
Db 961 TGACCGAGGTGGATGACACCATCTCAGCTCGGCTTGGCTGCCAATGGCTGGGTGTT 1020  
QY 1021 CTTGTTGGCTTATGTTAGTCCCGAGTTTGGCTGCTCACAAGCAACGAAACCCCATGGA 1080  
Db 1021 CTTGTTGGCTTATGTTAGTCCCGAGTTTGGCTGCTCACAAGCAACGAAACCCCATGGA 1080  
QY 1081 TTATCTGTTGAGGATGCTTCTGTAACCTCAACTCGTGAAGAGAGCTATGTTGTTGGA 1140  
Db 1081 TTATCTGTTGAGGATGCTTCTGTAACCTCAACTCGTGAAGAGAGCTATGTTGTTGGA 1140  
QY 1141 GAACAGAGCCTACTCTCAAGAGGAATCACTCAAGGTTTTTGAAGAGACAGGGACACGCT 1200  
Db 1141 GAACAGAGCCTACTCTCAAGAGGAATCACTCAAGGTTTTTGAAGAGACAGGGACACGCT 1200  
QY 1201 CTATGCCCTTATTCACACATTTTCAGCTGCAAGACCAAGCTCCCAAGAAATTTCTC 1260  
Db 1201 CTATGCCCTTATTCACACATTTTCAGCTGCAAGACCAAGCTCCCAAGAAATTTCTC 1260  
QY 1261 CATCCACGGGCCACGCTTGGCCGAGCCCTTCAAAAGACTATGAAGTAAAGAAAGGG 1320  
Db 1261 CATCCACGGGCCACGCTTGGCCGAGCCCTTCAAAAGACTATGAAGTAAAGAAAGGG 1320  
QY 1321 CAGCTAACTCTGCTGAAGAGTGGACAAATGCAGCCGGCGGCGAGATCTAGCGGGAGC 1380  
Db 1321 CAGCTAACTCTGCTGAAGAGTGGACAAATGCAGCCGGCGGCGAGATCTAGCGGGAGC 1380  
QY 1381 TCAAGGGATGTGGGGGAAATCTTGAGTCTTCTGAGAAACTGTACAAGACACTACGGGA 1440  
Db 1381 TCAAGGGATGTGGGGGAAATCTTGAGTCTTCTGAGAAACTGTACAAGACACTACGGGA 1440  
QY 1441 ACAGTTTGCTCCTCCAGCCTCAACCAATTTCTCATGCTGGGCTGATGGGCT 1500  
Db 1441 ACAGTTTGCTCCTCCAGCCTCAACCAATTTCTCATGCTGGGCTGATGGGCT 1500  
QY 1501 AGTAAGACTCCAGTCTTAGAGGCGCTGTAGTATTTTTTTTTTTTGTCTCATCCTTTGG 1560  
Db 1501 AGTAAGACTCCAGTCTTAGAGGCGCTGTAGTATTTTTTTTTTTTGTCTCATCCTTTGG 1560  
QY 1561 ATACTCTTTTAAAGTGGAGTCTCAGGCAACTCAAGTTTAGACCCCTTACTCTTTTGT 1620  
Db 1561 ATACTCTTTTAAAGTGGAGTCTCAGGCAACTCAAGTTTAGACCCCTTACTCTTTTGT 1620  
QY 1621 GTTTTTTGAACAGGATCTTGCTGTACCCAGGCTTGAGTCAGTGGTGCATCACAG 1680  
Db 1621 GTTTTTTGAACAGGATCTTGCTGTACCCAGGCTTGAGTCAGTGGTGCATCACAG 1680  
QY 1681 CCCAGTGCAGCCTCGACCCTGTGCTCAAGCAATCTCCCATCTCCATCTCCCAAGTG 1740  
Db 1681 CCCAGTGCAGCCTCGACCCTGTGCTCAAGCAATCTCCCATCTCCATCTCCCAAGTG 1740  
QY 1741 CTGGGATGACAGGCGTGAGCCACAGCTCCAGCCTAGGCCCTTAAATCTTGCTGTTATTT 1800  
Db 1741 CTGGGATGACAGGCGTGAGCCACAGCTCCAGCCTAGGCCCTTAAATCTTGCTGTTATTT 1800  
QY 1801 CCATGACTAAAGGCTGTGCTATCTGAGCTCAGCTGAGCTGCATCAGAGCTCTAGGGCCTG 1860  
Db 1801 CCATGACTAAAGGCTGTGCTATCTGAGCTCAGCTGAGCTGCATCAGAGCTCTAGGGCCTG 1860  
QY 1861 CTCCTCTAACTCAGTGGGTTTTGTGAGGCTCTGTGGCCACAGACAGCTCATATCT 1920  
Db 1861 CTCCTCTAACTCAGTGGGTTTTGTGAGGCTCTGTGGCCACAGACAGCTCATATCT 1920  
QY 1921 GAGCAAAAATAGCAAAAGCCTCTCTCAGCCCACTGGCCTGAAATCTACACTGGAAGCCAAC 1980  
Db 1921 GAGCAAAAATAGCAAAAGCCTCTCTCAGCCCACTGGCCTGAAATCTACACTGGAAGCCAAC 1980  
QY 1981 TTGCTGGCACCCCGCTCCCAACCTTCTTGCTGGTAGGAGAGGCTAAAGATCACCC 2040  
Db 1981 TTGCTGGCACCCCGCTCCCAACCTTCTTGCTGGTAGGAGAGGCTAAAGATCACCC 2040

QY 2041 TAAATTTACTCATCTCTAGTGTGCTGCCTCACAATGGGCTCAGAGCTCCCCAGACCA 2100  
Db 2041 TAAATTTACTCATCTCTAGTGTGCTGCCTCACAATGGGCTCAGAGCTCCCCAGACCA 2100  
QY 2101 ATTACAGGTCAACCCCTCTCTTCTTGCACTGTCCCAAACTTGTCTGTCATTCGAGATC 2160  
Db 2101 ATTACAGGTCAACCCCTCTCTTCTTGCACTGTCCCAAACTTGTCTGTCATTCGAGATC 2160  
QY 2161 TAAATCTCCCTTACGCTCTGCCAGGAATTTCTTTTTCAGACCTCCTAGCAGACGCGGTTG 2220  
Db 2161 TAAATCTCCCTTACGCTCTGCCAGGAATTTCTTTTTCAGACCTCCTAGCAGACGCGGTTG 2220  
QY 2221 CTCCTTGTGAGAGAAATTTGTAGATCATTTCTCACTTCAAATTCCTGGGCTGATCTTCT 2280  
Db 2221 CTCCTTGTGAGAGAAATTTGTAGATCATTTCTCACTTCAAATTCCTGGGCTGATCTTCT 2280  
QY 2281 CTCATCTTGCAACCCCACTCTGTAAATAGATTTACCGATTTACGGCTGATCTGTAA 2340  
Db 2281 CTCATCTTGCAACCCCACTCTGTAAATAGATTTACCGATTTACGGCTGATCTGTAA 2340  
QY 2341 GTGGGATGCTCTCAATAGGAGGAGTGTTCATTTGATATTAATAGTTATTACCTGAGTAT 2400  
Db 2341 GTGGGATGCTCTCAATAGGAGGAGTGTTCATTTGATATTAATAGTTATTACCTGAGTAT 2400  
QY 2401 GCAATAAAGATGTGGTGGCCACTCTTTTCATGTTGGTGGCAGCAAAAAA 2456  
Db 2401 GCAATAAAGATGTGGTGGCCACTCTTTTCATGTTGGTGGCAGCAAAAAA 2456

## RESULT 4

US-10-295-027-619  
; Sequence 619, Application US/10295027  
; Publication No. US200302350A1  
; GENERAL INFORMATION:  
; APPLICANT: Afar, Daniel  
; APPLICANT: Aziz, Natasha  
; APPLICANT: Ginsberg, Wendy M.  
; APPLICANT: Gish, Kurt C.  
; APPLICANT: Glynn, Richard  
; APPLICANT: Hevezi, Peter A.  
; APPLICANT: Mack, David H.  
; APPLICANT: Murray, Richard  
; APPLICANT: Watson, Susan R.  
; APPLICANT: Ros Biotechnology, Inc.  
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and  
; FILE REFERENCE: 018501-012500US  
; CURRENT APPLICATION NUMBER: US/10/295,027  
; CURRENT FILING DATE: 2002-11-13  
; PRIOR APPLICATION NUMBER: US 09/663,733  
; PRIOR FILING DATE: 2000-09-15  
; PRIOR APPLICATION NUMBER: US 60/350,666  
; PRIOR FILING DATE: 2001-11-13  
; PRIOR APPLICATION NUMBER: US 60/335,394  
; PRIOR FILING DATE: 2001-11-15  
; PRIOR APPLICATION NUMBER: US 60/332,464  
; PRIOR FILING DATE: 2001-11-21  
; PRIOR APPLICATION NUMBER: US 60/334,393  
; PRIOR FILING DATE: 2001-11-29  
; PRIOR APPLICATION NUMBER: US 60/340,376  
; PRIOR FILING DATE: 2001-12-14  
; PRIOR APPLICATION NUMBER: US 60/347,211  
; PRIOR FILING DATE: 2002-01-08  
; PRIOR APPLICATION NUMBER: US 60/347,349  
; PRIOR FILING DATE: 2002-01-10  
; PRIOR APPLICATION NUMBER: US 60/355,250  
; PRIOR FILING DATE: 2002-02-08  
; PRIOR APPLICATION NUMBER: US 60/356,714  
; PRIOR FILING DATE: 2002-02-13  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 1386  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 619

; LENGTH: 2456			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
US-10-295-027-619			
Query Match			
Best Local Similarity			
Matches 2456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	1	ATAACAGCATGAAGTCCGCTGGAACTGGAAATAGGCGGTCTCTCCCTCGACCCCTCCCC	60
DB	1	ATPACAGCATGAAGTCCGCTGGAACTGGAAATAGGCGGTCTCTCCCTCGACCCCTCCCC	60
QY	61	TCTTGTCCCTCTGCTCACCCCTCGCTTCCCTCCCTCCGCGAGGCCCTTTATA	120
DB	61	TCTTGTCCCTCTGCTCACCCCTCGCTTCCCTCCCTCCGCGAGGCCCTTTATA	120
QY	121	ACAACTGCTCAGATGCGAGGGGGGATAGCTGTCCAAGGTCTCCCGCAGCACTGAGGAG	180
DB	121	ACAACTGCTCAGATGCGAGGGGGGATAGCTGTCCAAGGTCTCCCGCAGCACTGAGGAG	180
QY	181	CTCGCTGCTGCCCTCTTCGCGCGGGGAAGCAGCACCAAGTTTCACGGCCAAAGCCCTTGGC	240
DB	181	CTCGCTGCTGCCCTCTTCGCGCGGGGAAGCAGCACCAAGTTTCACGGCCAAAGCCCTTGGC	240
QY	241	ACTAGGGTCCAGATGGCTTACAACTGCTCCCTGATGGTTCCGCAATGGCTGAATCCAA	300
DB	241	ACTAGGGTCCAGATGGCTTACAACTGCTCCCTGATGGTTCCGCAATGGCTGAATCCAA	300
QY	301	GTACTACAGACTTTGTGATAAGGCTGAAGCTTGGGGCATCGTCTAGAAACGGTGGCCAC	360
DB	301	GTACTACAGACTTTGTGATAAGGCTGAAGCTTGGGGCATCGTCTAGAAACGGTGGCCAC	360
QY	361	AGCCGGGTTGTGACCTCGGTGGCCCTTATGCTCACTCTCCGATCCTCGTCTGCAAGGT	420
DB	361	AGCCGGGTTGTGACCTCGGTGGCCCTTATGCTCACTCTCCGATCCTCGTCTGCAAGGT	420
QY	421	GCAGGACTCCAAAGCGGAAAAATGCTGCTACTCAGATTTCTTCTCCTCGGTGTGTT	480
DB	421	GCAGGACTCCAAAGCGGAAAAATGCTGCTACTCAGATTTCTTCTCCTCGGTGTGTT	480
QY	481	GGGCATCTTTGGCCTCACCTTCGCTTCACTCGGACTGGAGCGGAGCACAGGGCCAC	540
DB	481	GGGCATCTTTGGCCTCACCTTCGCTTCACTCGGACTGGAGCGGAGCACAGGGCCAC	540
QY	541	ACGCTTCTTCTCTTTGGGATCCTCTTTTCCATCTGCTTCTCTGCTGCTGGCTCATGC	600
DB	541	ACGCTTCTTCTCTTTGGGATCCTCTTTTCCATCTGCTTCTCTGCTGCTGGCTCATGC	600
QY	601	TGTCAGTCTGACCAAGCTCGTCCGGGGAGGAAGCCCTTTCCCTGTTGGTGAATCTGGG	660
DB	601	TGTCAGTCTGACCAAGCTCGTCCGGGGAGGAAGCCCTTTCCCTGTTGGTGAATCTGGG	660
QY	661	TCGCGCGTGGGCTTCAGCCTAGTCAGAGATGTTATCGCTATTGAATATTGTCCTGAC	720
DB	661	TCGCGCGTGGGCTTCAGCCTAGTCAGAGATGTTATCGCTATTGAATATTGTCCTGAC	720
QY	721	CATGAATAGGACCAACGCTCAATGTCTTTTCTGAGCTTTCCGCTCTCGTCGCAATGAAGA	780
DB	721	CATGAATAGGACCAACGCTCAATGTCTTTTCTGAGCTTTCCGCTCTCGTCGCAATGAAGA	780
QY	781	CTTTGTCTCTGCTCACTAGTCCTCTTCTTTGATGGCGTGAACCTTCTCATGTCTCTC	840
DB	781	CTTTGTCTCTGCTCACTAGTCCTCTTCTTTGATGGCGTGAACCTTCTCATGTCTCTC	840
QY	841	CTTCACTTCTGTTGTTCTTCAAGGCTGGAAGACATGGGGCCACATCTACTCTAC	900
DB	841	CTTCACTTCTGTTGTTCTTCAAGGCTGGAAGACATGGGGCCACATCTACTCTAC	900
QY	901	GATGCTCTCTCCATTTGCCATCTGGGTGGCCTGGATCACCCCTGCTCATGCTTCTGACTT	960
DB	901	GATGCTCTCTCCATTTGCCATCTGGGTGGCCTGGATCACCCCTGCTCATGCTTCTGACTT	960

QY	961	TGACCGCAGGTGGGATGACACCATCTCTAGCTCCGGCTTGGCTGCCAATGGCTGGGTGTT	1020
DB	961	TGACCGCAGGTGGGATGACACCATCTCTAGCTCCGGCTTGGCTGCCAATGGCTGGGTGTT	1020
QY	1021	CCTGTTGGCTTATGTAGTCCCGAGTTTGGCTGTCTCAAAAGCAACGAAACCCCATGGA	1080
DB	1021	CCTGTTGGCTTATGTAGTCCCGAGTTTGGCTGTCTCAAAAGCAACGAAACCCCATGGA	1080
QY	1081	TTATCTGTTGAGGATGCTTTCTGTAAACCTCAACTCGTGAAGAGAGCTATGTTGTGA	1140
DB	1081	TTATCTGTTGAGGATGCTTTCTGTAAACCTCAACTCGTGAAGAGAGCTATGTTGTGA	1140
QY	1141	GAAACAGCCTACTCTCAAGAGAAATCACTCAAGGTTTGAAGAGAGAGCTATGTTGTGA	1200
DB	1141	GAAACAGCCTACTCTCAAGAGAAATCACTCAAGGTTTGAAGAGAGAGCTATGTTGTGA	1200
QY	1201	CTATGCCCTTATTCACACATTTTTCAGCTGAGAACCAAGCCTCCCAAAAGAAATTTCTC	1260
DB	1201	CTATGCCCTTATTCACACATTTTTCAGCTGAGAACCAAGCCTCCCAAAAGAAATTTCTC	1260
QY	1261	CATCCACAGGGCCCGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAGAAAGAGGG	1320
DB	1261	CATCCACAGGGCCCGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAGAAAGAGGG	1320
QY	1321	CAGCTAACTCTGCTGAAGAGTGGGACAAATGCGCCGGCGGCGAGATCTAGCGGGAGC	1380
DB	1321	CAGCTAACTCTGCTGAAGAGTGGGACAAATGCGCCGGCGGCGAGATCTAGCGGGAGC	1380
QY	1381	TCAAAAGGATGCGGCAAAATCTTGAGTCTTCTGAGAAACTGTACAACACTACCGGA	1440
DB	1381	TCAAAAGGATGCGGCAAAATCTTGAGTCTTCTGAGAAACTGTACAACACTACCGGA	1440
QY	1441	ACAGTTTGCCTCCCTCCAGCCTCAACCAAAATTTCTTCATGCTGGGCTGATGGGCT	1500
DB	1441	ACAGTTTGCCTCCCTCCAGCCTCAACCAAAATTTCTTCATGCTGGGCTGATGGGCT	1500
QY	1501	AGTAAGACTCAGTTCTTAGAGGCGCTGATGATTTTTTTTTTTTGTCTCATCTCTTGG	1560
DB	1501	AGTAAGACTCAGTTCTTAGAGGCGCTGATGATTTTTTTTTTTTGTCTCATCTCTTGG	1560
QY	1561	ATACTTCTTTTAAAGTGGGAGTCTCAGGCAACTCAAGATTTAGACCCCTTACTCTTTTGT	1620
DB	1561	ATACTTCTTTTAAAGTGGGAGTCTCAGGCAACTCAAGATTTAGACCCCTTACTCTTTTGT	1620
QY	1621	GTTTTTTGAACAGAGATCTTGTCTGTCAACCAGGCTTGAGTGCAAGTGTGCGATCACAG	1680
DB	1621	GTTTTTTGAACAGAGATCTTGTCTGTCAACCAGGCTTGAGTGCAAGTGTGCGATCACAG	1680
QY	1681	CCAGTGCAGCCTCGAACCCCTGCTCAGCAATCCTCCCATCTCCATCTCCCAAGTG	1740
DB	1681	CCAGTGCAGCCTCGAACCCCTGCTCAGCAATCCTCCCATCTCCATCTCCCAAGTG	1740
QY	1741	CTGGGATGACAGGCTGAGCCACAGCTCCAGCCTAGGCCCTTAATCTTGTGTTATTTT	1800
DB	1741	CTGGGATGACAGGCTGAGCCACAGCTCCAGCCTAGGCCCTTAATCTTGTGTTATTTT	1800
QY	1801	CCATGAGCTAAAGGTCTGGTCACTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCT	1860
DB	1801	CCATGAGCTAAAGGTCTGGTCACTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCT	1860
QY	1861	CTCCTCTAACTCAGATGGGTTTGTGAGGCTCTGTGGCCAGAGACAGCTGCATATCT	1920
DB	1861	CTCCTCTAACTCAGATGGGTTTGTGAGGCTCTGTGGCCAGAGACAGCTGCATATCT	1920
QY	1921	GAGCAAAATAGCAAAAGCCTCTCTCAGCCCACTGGCCTGAAATCTACACTGGAAGCAAC	1980
DB	1921	GAGCAAAATAGCAAAAGCCTCTCTCAGCCCACTGGCCTGAAATCTACACTGGAAGCAAC	1980
QY	1981	TTGCTGGGACCCCGCTCCCAACCCCTTCTTGGCTGGGTAGGAGGCTTAAAGATCACCC	2040
DB	1981	TTGCTGGGACCCCGCTCCCAACCCCTTCTTGGCTGGGTAGGAGGCTTAAAGATCACCC	2040
QY	2041	TAAATTTACTCATCTCTTAGTGCTCTACATTTGGGCTCAGCAGCTCCCGAGCACA	2100

```
Db 2041 TAAATTTACTCATCTCTAGTCTGCTGCCTCACATTTGGGCTCAGAGCTCCCCAGACCA 2100
Qy 2101 ATTACAGGTACCCCTCTCTCTTCTTGCACTGTCGCCAAACTTGCTGTAATCCGAGATC 2160
Db 2101 ATTACAGGTACCCCTCTCTCTTCTTGCACTGTCGCCAAACTTGCTGTAATCCGAGATC 2160
Qy 2161 TAATCTCCCCCTACGCTCTGCCAGGAATCTTTTCAGACCTCACTAGCACAAAGCCGGTTG 2220
Db 2161 TAATCTCCCCCTACGCTCTGCCAGGAATCTTTTCAGACCTCACTAGCACAAAGCCGGTTG 2220
Qy 2221 CTCCTTGTGAGGAGAAATTTGTAGATCATTTCTCACTTCAAATTCCTGGGCTGATCTTCT 2280
Db 2221 CTCCTTGTGAGGAGAAATTTGTAGATCATTTCTCACTTCAAATTCCTGGGCTGATCTTCT 2280
Qy 2281 CTCATCTTGACCCCAACCTCTGTAATAGATTTACCGCATTTTACGGCTGCAATCTGTAA 2340
Db 2281 CTCATCTTGACCCCAACCTCTGTAATAGATTTACCGCATTTTACGGCTGCAATCTGTAA 2340
Qy 2341 GTGGGATGCTCTCCTAAATGGAGAGTGTTCATTGTATAATAGTTATTACCTGAGTAT 2400
Db 2341 GTGGGATGCTCTCCTAAATGGAGAGTGTTCATTGTATAATAGTTATTACCTGAGTAT 2400
Qy 2401 GCAATAAAGATGTGTGGCCACTCTTTTCATGTGTGGTGGCAGCAAAAAA 2456
Db 2401 GCAATAAAGATGTGTGGCCACTCTTTTCATGTGTGGTGGCAGCAAAAAA 2456

RESULT 5
US-10-600-816-2
; Sequence 2, Application US/10600816
; Publication No. US20040121362A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: IDENTIFICATION AND MODULATION OF A G-PROTEIN COUPLED RECEPTOR
; TITLE OF INVENTION: (GPCR), RA13, ASSOCIATED WITH CHRONIC OBSTRUCTIVE PULMONARY
; TITLE OF INVENTION: DISEASE (COPD) AND NF-KB AND E-SELECTIN REGULATION
; FILE REFERENCE: D0251 NP
; CURRENT APPLICATION NUMBER: US/10/600,816
; CURRENT FILING DATE: 2003-06-20
; PRIOR APPLICATION NUMBER: U.S. 60/390,850
; PRIOR FILING DATE: 2002-06-20
; PRIOR APPLICATION NUMBER: U.S. 60/407,006
; PRIOR FILING DATE: 2002-08-29
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 2456
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-600-816-2

Query Match 100.0%; Score 2456; DB 8; Length 2456;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATACAGCATGAAGTCCGTGGAACTGGAATAGCGGTGCTCTCCCTCGACCCCTCCCC 60
Db 1 ATAACAGCATGAAGTCCGTGGAACTGGAATAGCGGTGCTCTCCCTCGACCCCTCCCC 60
Qy 61 TCCTTGCTCCTCTGCTCACCCCTCGCTCGTTCCCTCCCGGAGGGCCGCTTTATA 120
Db 61 TCCTTGCTCCTCTGCTCACCCCTCGCTCGTTCCCTCCCGGAGGGCCGCTTTATA 120
Qy 121 ACAATGCTCAGAGTCGAGGGCGGATAGCTGTCCAAGTCTCCCCAGCACTAGGAG 180
Db 121 ACAATGCTCAGAGTCGAGGGCGGATAGCTGTCCAAGTCTCCCCAGCACTAGGAG 180
Qy 181 CTCGCTGCTGCCCTCTTGGCGGGAGGACGACCAAGTTTCAACGCCCAACGCTTGGC 240
Db 181 CTCGCTGCTGCCCTCTTGGCGGGAGGACGACCAAGTTTCAACGCCCAACGCTTGGC 240
Qy 241 ACTAGGCTCCAGAATGGCTACAACTGCTCCCTGATGGTTGGCCGAATGGGCTGAATCCAA 300
```

```
Db 241 ACTAGGCTCCAGAATGGCTACAACTGCTCCCTGATGGTTGGCCGAATGGGCTGAATCCAA 300
Qy 301 GTACTACAGACTTTGTGATAAGCGTGAAGCTTGGGGCATCGTCTAGAAACCGTGGCCAC 360
Db 301 GTACTACAGACTTTGTGATAAGCGTGAAGCTTGGGGCATCGTCTAGAAACCGTGGCCAC 360
Qy 361 AGCCGGGGTTGTGACCTCGGTGGCCCTTCAATGCTCACTCTCCGATCCTCGTCTGCAAGGT 420
Db 361 AGCCGGGGTTGTGACCTCGGTGGCCCTTCAATGCTCACTCTCCGATCCTCGTCTGCAAGGT 420
Qy 421 GCAGGACTCCAA CAGSCGAAAAATGCTCCTACTCAAGTTTCTTCTTCTCTCTGGGTGTGTT 480
Db 421 GCAGGACTCCAA CAGSCGAAAAATGCTCCTACTCAAGTTTCTTCTTCTCTGGGTGTGTT 480
Qy 481 GGGCATCTTTGGCCCTCACCTTCGCTTTCATCATCGGACTGGAGCGGAGCACAGGGCCAC 540
Db 481 GGGCATCTTTGGCCCTCACCTTCGCTTTCATCATCGGACTGGAGCGGAGCACAGGGCCAC 540
Qy 541 AGCTTCTTCTCTTTGGGATCCTCTTTTCCATCTGCTTCTCTGCTCTGCTGCTCATGC 600
Db 541 AGCTTCTTCTCTTTGGGATCCTCTTTTCCATCTGCTTCTCTGCTCTGCTGCTCATGC 600
Qy 601 TGTCACTGTGACCAAGCTCGTCCGGGGAGGAAGCCCTTTTCCCTGTTGGTGAATTCGGG 660
Db 601 TGTCACTGTGACCAAGCTCGTCCGGGGAGGAAGCCCTTTTCCCTGTTGGTGAATTCGGG 660
Qy 661 TCTGGCGTGGGCTTCAGCCTAGTCAGGATGTTATCGCTATTTGAATATATATGTCCTGAC 720
Db 661 TCTGGCGTGGGCTTCAGCCTAGTCAGGATGTTATCGCTATTTGAATATATATGTCCTGAC 720
Qy 721 CATGAATAGGACCAACGTCATGTCCTTTCTGAGCTTTCGGCTCTCTCGCAATGAAGA 780
Db 721 CATGAATAGGACCAACGTCATGTCCTTTCTGAGCTTTCGGCTCTCTCGCAATGAAGA 780
Qy 781 CTTTGTCTCTCTCACTACCTGCTCTTCTTTGATGGCGCTGACCTTCTCTCATGTCCTC 840
Db 781 CTTTGTCTCTCTCACTACCTGCTCTTCTTTGATGGCGCTGACCTTCTCTCATGTCCTC 840
Qy 841 CTTACCTTCTGCTGTTTCTTCCACGGGCTGGAGAGACATGGGGCCACATCTACTCTAC 900
Db 841 CTTACCTTCTGCTGTTTCTTCCACGGGCTGGAGAGACATGGGGCCACATCTACTCTAC 900
Qy 901 GATGCTCTCTCCATTTGCCATCTGGGTGGCTGGATCACCTGCTCATGCTCTCTGACTT 960
Db 901 GATGCTCTCTCCATTTGCCATCTGGGTGGCTGGATCACCTGCTCATGCTCTCTGACTT 960
Qy 961 TGACCGCAGGTGGGATGACACCATCTCAGCTCCGCTTTGGCTGCAATGGGTGGGTGTT 1020
Db 961 TGACCGCAGGTGGGATGACACCATCTCAGCTCCGCTTTGGCTGCAATGGGTGGGTGTT 1020
Qy 1021 CTTGTTGGCTTATGTTAGTCCGAGTTTGGCTGCTCAAAAGCAACGAAACCCCATGGA 1080
Db 1021 CTTGTTGGCTTATGTTAGTCCGAGTTTGGCTGCTCAAAAGCAACGAAACCCCATGGA 1080
Qy 1081 TTATCTCTGTTGAGGATGCTTTCTGTAACCTCAAACCTGTAAGAGAGCTATGGTGTGGA 1140
Db 1081 TTATCTCTGTTGAGGATGCTTTCTGTAACCTCAAACCTGTAAGAGAGCTATGGTGTGGA 1140
Qy 1141 GAACGAGCCTACTCTCAAGAGGAAATCACTCAAGGTTTTGAAGAGACAGGGACACGCT 1200
Db 1141 GAACGAGCCTACTCTCAAGAGGAAATCACTCAAGGTTTTGAAGAGACAGGGACACGCT 1200
Qy 1201 CTATGCCCCCTATTCCACATTTTTCAGCTCGAAGACAGCCTCCCAAGAAAGAAATTC 1260
Db 1201 CTATGCCCCCTATTCCACATTTTTCAGCTCGAAGACAGCCTCCCAAGAAAGAAATTC 1260
Qy 1261 CATCCCAAGGGCCACGCTTGGCCGAGCCTTACAAAGACTATGAAGTAAAGAAAGAGG 1320
Db 1261 CATCCCAAGGGCCACGCTTGGCCGAGCCTTACAAAGACTATGAAGTAAAGAAAGAGG 1320
Qy 1321 CAGCTAACTCTGCTGTAAGAGTGGGACAAATTCAGCCGGGGCGGAGATCTTAGCGGAGC 1380
Db 1321 CAGCTAACTCTGCTGTAAGAGTGGGACAAATTCAGCCGGGGCGGAGATCTTAGCGGAGC 1380
```



Db 661 TCTGGCGGTGGGCTTACGCTAGTCAGGATGTTATCGCTATTGAATATATGTCCTGAC 720  
Qy 721 CATGAATAGAACCAACGTCATGCTTCTTCTGAGCTTTCCGCTCCGTCGCAATGAAGA 780  
Db 721 CATGAATAGAACCAACGTCATGCTTCTTCTGAGCTTTCCGCTCCGTCGCAATGAAGA 780  
Qy 781 CTTTGTCTCTCTGCTCAGCTACGCTCCTCTCTGATGGGCTGACCTTCTCTCATGTCCTC 840  
Db 781 CTTTGTCTCTCTGCTCAGCTACGCTCCTCTCTGATGGGCTGACCTTCTCTCATGTCCTC 840  
Qy 841 CTTTACGCTTCTGCTGCTTCTTCAAGGCTGGAAGAGACATGGGGCCCAATCTACCTCAC 900  
Db 841 CTTTACGCTTCTGCTGCTTCTTCAAGGCTGGAAGAGACATGGGGCCCAATCTACCTCAC 900  
Qy 901 GATGCTCTCTCATTGCGCATCTGGGTGGCTGGATCACCTGCTCATGCTTCTGACTT 960  
Db 901 GATGCTCTCTCATTGCGCATCTGGGTGGCTGGATCACCTGCTCATGCTTCTGACTT 960  
Qy 961 TGACCCGAGTGGGATGACACCATCTCAGCTCCGCTTGGCTGCCAATGGCTGGGTGTT 1020  
Db 961 TGACCCGAGTGGGATGACACCATCTCAGCTCCGCTTGGCTGCCAATGGCTGGGTGTT 1020  
Qy 1021 CTTGTTGGCTTATGTTAGTCCCGAGTTTGGCTGCTCAGAAAGCAACCCCATGGA 1080  
Db 1021 CTTGTTGGCTTATGTTAGTCCCGAGTTTGGCTGCTCAGAAAGCAACCCCATGGA 1080  
Qy 1081 TTATCTCTGTTGAGGATGCTTTCTGTAACCTCACTGCTGGAAGAGACTATGGTGTGA 1140  
Db 1081 TTATCTCTGTTGAGGATGCTTTCTGTAACCTCACTGCTGGAAGAGACTATGGTGTGA 1140  
Qy 1141 GAACGAGCTACTCTCAGAGAAATCACTCAAGCTTTTGAAGACAGGGGACACGCT 1200  
Db 1141 GAACGAGCTACTCTCAGAGAAATCACTCAAGCTTTTGAAGACAGGGGACACGCT 1200  
Qy 1201 CTATGCCCCCTATTTCACACATTTTTCAGCTGAGAAACCAAGCTTCCCAAGAAATTTCTC 1260  
Db 1201 CTATGCCCCCTATTTCACACATTTTTCAGCTGAGAAACCAAGCTTCCCAAGAAATTTCTC 1260  
Qy 1261 CATCCACAGGGCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAGAAAGAGGG 1320  
Db 1261 CATCCACAGGGCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAGAAAGAGGG 1320  
Qy 1321 CAGCTAACTCTGCTCAGAGTGGGACAAATGACCGCGGCGGAGACTACGGGAGC 1380  
Db 1321 CAGCTAACTCTGCTCAGAGTGGGACAAATGACCGCGGCGGAGACTACGGGAGC 1380  
Qy 1381 TCAAGGGATGTGGGCGAAATCTTGAGTCTTCTGAGAAACTGTACAAGACACTACGGGA 1440  
Db 1381 TCAAGGGATGTGGGCGAAATCTTGAGTCTTCTGAGAAACTGTACAAGACACTACGGGA 1440  
Qy 1441 ACAGTTTGCCTCCCTCCAGGCTCAACCAATTTCCATGCTGGGGCTGATGGGCT 1500  
Db 1441 ACAGTTTGCCTCCCTCCAGGCTCAACCAATTTCCATGCTGGGGCTGATGGGCT 1500  
Qy 1501 AGTAAGACTCCAGTCTTGAAGCGGCTGTAGTATTTTCTTTTGTCTCATCCTTTGG 1560  
Db 1501 AGTAAGACTCCAGTCTTGAAGCGGCTGTAGTATTTTCTTTTGTCTCATCCTTTGG 1560  
Qy 1561 ATACTTCTTTTAAAGTGGGAGTCTCAGGCAACTCAAGTTTGAAGCCCTTACTCTTTTGT 1620  
Db 1561 ATACTTCTTTTAAAGTGGGAGTCTCAGGCAACTCAAGTTTGAAGCCCTTACTCTTTTGT 1620  
Qy 1621 GTTTTTTGAACAGGATCTTCTGTACCCAGGCTTGAAGTCAAGTGGGATCACAG 1680  
Db 1621 GTTTTTTGAACAGGATCTTCTGTACCCAGGCTTGAAGTCAAGTGGGATCACAG 1680  
Qy 1681 CCCAGTGCAGCTCGACCACTGCTGCTCAGCAATCTCCATCTCCATCTCCCAAGTG 1740  
Db 1681 CCCAGTGCAGCTCGACCACTGCTGCTCAGCAATCTCCATCTCCATCTCCCAAGTG 1740  
Qy 1741 CTGGGATGACAGGCTGAGCCACAGCTCCAGGCTTAGGCCCTTAAATCTTCTGCTGTTT 1800

Db 1741 CTGGGATGACAGGCGTAGCCACAGCTCCAGCTCCAGCTTAGGCCCTTAAATCTTGTGTTATTT 1800  
Qy 1801 CMATGAGCTAAAGGTCTGTGTCATCTGAGCTCAAGCTGGCTCAACAGCTCTTAGGGGCTG 1860  
Db 1801 CMATGAGCTAAAGGTCTGTGTCATCTGAGCTCAAGCTGGCTCAACAGCTCTTAGGGGCTG 1860  
Qy 1861 CTCCTCTAACTACAGTGGGTTTGTGAGGCTCTGTGGCCAGAGAGAGCTGTCATATCT 1920  
Db 1861 CTCCTCTAACTACAGTGGGTTTGTGAGGCTCTGTGGCCAGAGAGAGCTGTCATATCT 1920  
Qy 1921 GAGCAAAATAGCAAAAGGCTCTCTCAGCCCACTGGCTGAAATCTACACTGGAAGCAAC 1980  
Db 1921 GAGCAAAATAGCAAAAGGCTCTCTCAGCCCACTGGCTGAAATCTACACTGGAAGCAAC 1980  
Qy 1981 TTGCTGGCACCCCGCTCCCAACCTTCTTGGCTGGTAGGAGAGCTAAAGATCACCC 2040  
Db 1981 TTGCTGGCACCCCGCTCCCAACCTTCTTGGCTGGTAGGAGAGCTAAAGATCACCC 2040  
Qy 2041 TAAATTTACTCATCTCTCTAGTGTCTGCTCAGCTTGGGCTCAGAGCTCCCGAGACCA 2100  
Db 2041 TAAATTTACTCATCTCTCTAGTGTCTGCTCAGCTTGGGCTCAGAGCTCCCGAGACCA 2100  
Qy 2101 ATTACAGGTGACCCCTCTCTTCTGCACTGTCCCAAACTTGTCTGCAATTCGAGATC 2160  
Db 2101 ATTACAGGTGACCCCTCTCTTCTGCACTGTCCCAAACTTGTCTGCAATTCGAGATC 2160  
Qy 2161 TAAATCTCCCTTACGCTCTGCGAGGAATTTCTTTCAGACCTCAGTACAGAGCCCGGTTG 2220  
Db 2161 TAAATCTCCCTTACGCTCTGCGAGGAATTTCTTTCAGACCTCAGTACAGAGCCCGGTTG 2220  
Qy 2221 CTCCTTGTGAGGAAATTTGTAGATCATTTCTCACTTCAATTTCTGGGCTGATCTTCT 2280  
Db 2221 CTCCTTGTGAGGAAATTTGTAGATCATTTCTCACTTCAATTTCTGGGCTGATCTTCT 2280  
Qy 2281 CTCATCTTGACCCCAACCTCTGTAATAGATTTACCGCATTTACGGCTGATTTGTAA 2340  
Db 2281 CTCATCTTGACCCCAACCTCTGTAATAGATTTACCGCATTTACGGCTGATTTGTAA 2340  
Qy 2341 GTGGGATGGTCTCTTAATGGAGAGTGTCTATGATTAATAGTTATTCACCTGAGTAT 2400  
Db 2341 GTGGGATGGTCTCTTAATGGAGAGTGTCTATGATTAATAGTTATTCACCTGAGTAT 2400  
Qy 2401 GCAATAAAGATGTGGGCACTCTTCTATGTTGGTGGGCAAGCAAAAAA 2456  
Db 2401 GCAATAAAGATGTGGGCACTCTTCTATGTTGGTGGGCAAGCAAAAAA 2456

RESULT 7

US-10-936-626-40  
; Sequence 40, Application US/10936626  
; Publication No. US20050106644A1  
; GENERAL INFORMATION:  
; APPLICANT: Cairns, Belinda  
; APPLICANT: Chen, Ruihuan  
; APPLICANT: Frantz, Gretchen  
; APPLICANT: Hillan, Kenneth J.  
; APPLICANT: Koeppe, Hartmut  
; APPLICANT: Phillips, Heidi S.  
; APPLICANT: Polakis, Paul  
; APPLICANT: Spencer, Susan D.  
; APPLICANT: Smith, Victoria  
; APPLICANT: Williams, P. Mickey  
; APPLICANT: Wu, Thomas D.  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: Compositions and Methods for the Diagnosis and  
; TITLE OF INVENTION: Treatment of Tumor  
; FILE REFERENCE: P5001R1P1  
; CURRENT APPLICATION NUMBER: US/10/936,626  
; CURRENT FILING DATE: 2004-09-08  
; PRIOR APPLICATION NUMBER: US 10/872,991  
; PRIOR FILING DATE: 2004-06-21  
; PRIOR APPLICATION NUMBER: US 10/872,972  
; PRIOR FILING DATE: 2004-06-21



; PRIOR APPLICATION NUMBER: US 10/241,220
; PRIOR FILING DATE: 2002-09-11
; PRIOR APPLICATION NUMBER: US 10/177,488
; PRIOR FILING DATE: 2002-06-19
; PRIOR APPLICATION NUMBER: US 60/299,500
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 60/301,880
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/323,268
; PRIOR FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US 60/557,116
; PRIOR FILING DATE: 2004-03-26
; PRIOR APPLICATION NUMBER: US 60/598,899
; PRIOR FILING DATE: 2004-08-04
; NUMBER OF SEQ ID NOS: 154
; SEQ ID NO 40
; LENGTH: 2456
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-936-626-40

Query Match
Best Local Similarity 100.0%; Score 2456; DB 10; Length 2456;
Matches 2456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATACAGCATGAAGTGCCTGGAACTGGAATAGGGGTGTCTCTCCCTCGACCCCTCCCCC 60
DB 1 ATACAGCATGAAGTGCCTGGAACTGGAATAGGGGTGTCTCTCCCTCGACCCCTCCCCC 60

QY 61 TCTTTGTCCTCTGTCAACCCCTCGCTCGTTCCTCCCTCCCTCCGGGAGGGCCGCTTTATA 120
DB 61 TCTTTGTCCTCTGTCAACCCCTCGCTCGTTCCTCCCTCCCTCCGGGAGGGCCGCTTTATA 120

QY 121 ACAACTGCTCAGAGTGCAGGGCGGGATAGCTGTCCAAGGTCTCCCCAGCACTGAGGAG 180
DB 121 ACAACTGCTCAGAGTGCAGGGCGGGATAGCTGTCCAAGGTCTCCCCAGCACTGAGGAG 180

QY 181 CTGCGCTGTGCTTCTTGGCGCGGGAAAGCAGCACCAGATTTACGGCCAAACGCTTTGGC 240
DB 181 CTGCGCTGTGCTTCTTGGCGCGGGAAAGCAGCACCAGATTTACGGCCAAACGCTTTGGC 240

QY 241 ACTAGGTTCCAGATGGCTTACACAGTCCCTCATGGTTGCGCCCAATGGCTGAATCCAA 300
DB 241 ACTAGGTTCCAGATGGCTTACACAGTCCCTCATGGTTGCGCCCAATGGCTGAATCCAA 300

QY 301 GTACTACAGACTTTGTGATAAGCTGAAGCTTTGGGGCATCGTCTAGAAAACGGTGGCCAC 360
DB 301 GTACTACAGACTTTGTGATAAGCTGAAGCTTTGGGGCATCGTCTAGAAAACGGTGGCCAC 360

QY 361 AGCCGGGGTTGTGACCTCGGTGGCCCTTCATGTCTCACTCTCCGGATCCTCGTCTGCAAGGT 420
DB 361 AGCCGGGGTTGTGACCTCGGTGGCCCTTCATGTCTCACTCTCCGGATCCTCGTCTGCAAGGT 420

QY 421 GCAGACTTCAACAGGCGAAAATGCTGCTACTCAAGTTTCTTCTTCTCTCGGTGTGTT 480
DB 421 GCAGACTTCAACAGGCGAAAATGCTGCTACTCAAGTTTCTTCTTCTCTCGGTGTGTT 480

QY 481 GGGCATCTTTGGCTCACCTCGCTTCATCATCGGACTGGACGGGAGCACAGGGGCCAC 540
DB 481 GGGCATCTTTGGCTCACCTCGCTTCATCATCGGACTGGACGGGAGCACAGGGGCCAC 540

QY 541 ACGCTTCTTCTCTTTGGGATCCTCTTTTCCATCTGCTTTCTCTCTGCTGCTGGCTCATGC 600
DB 541 ACGCTTCTTCTCTTTGGGATCCTCTTTTCCATCTGCTTTCTCTCTGCTGCTGGCTCATGC 600

QY 601 TGTGAGTCTGACCAAGCTCGTCCGGGGAGGAGCCCTTTCCCTGTTGCTGATCTGGG 660
DB 601 TGTGAGTCTGACCAAGCTCGTCCGGGGAGGAGCCCTTTCCCTGTTGCTGATCTGGG 660

QY 661 TCTGGCCGTGGGCTTCAAGCTTAGTCAGGATGTTATCGCTATTGAATATATTGCTCTGAC 720
DB 661 TCTGGCCGTGGGCTTCAAGCTTAGTCAGGATGTTATCGCTATTGAATATATTGCTCTGAC 720

QY 721 CATGAATAGGACCAACGTCATGTCTTTTCTGAGCTTTCCGCTCTCTCGTCCAAATGAAGA 780
DB CATGAATAGGACCAACGTCATGTCTTTTCTGAGCTTTCCGCTCTCTCGTCCAAATGAAGA 780

QY 781 CTTTGTCTCTCGCTCACCTACGTCTCTCTTTGATGGCGTGACCTTCTCATGTCCCTC 840
DB CTTTGTCTCTCGCTCACCTACGTCTCTCTTTGATGGCGTGACCTTCTCATGTCCCTC 840

QY 841 CTTTCACTTCTGTGGTTCTTTCACGGGCTGGAAGAGACATGGGGCCACATCTACCTCAC 900
DB CTTTCACTTCTGTGGTTCTTTCACGGGCTGGAAGAGACATGGGGCCACATCTACCTCAC 900

QY 901 GATGCTCTCTCCATTTGCCATCTGGGTGSCCTGGATCACCTTCTCATGCTTCTCTGACTT 960
DB GATGCTCTCTCCATTTGCCATCTGGGTGSCCTGGATCACCTTCTCATGCTTCTCTGACTT 960

QY 961 TGACCGCAGGTGGGATGACACCATCTCAGCTCCGCTTGGCTGCGCAATGGCTGGGTGTT 1020
DB TGACCGCAGGTGGGATGACACCATCTCAGCTCCGCTTGGCTGCGCAATGGCTGGGTGTT 1020

QY 1021 CTTGTTGGCTTATGTTAGTCCGAGTTTGGCTGTCTCAAAAGCAACGAAACCCCATGGA 1080
DB CTTGTTGGCTTATGTTAGTCCGAGTTTGGCTGTCTCAAAAGCAACGAAACCCCATGGA 1080

QY 1081 TTATCCTGTTGAGGATGCTTTCTGTAAACCTCAACTCTGTTGAAGAGAGCTATGGTGGGA 1140
DB TTATCCTGTTGAGGATGCTTTCTGTAAACCTCAACTCTGTTGAAGAGAGCTATGGTGGGA 1140

QY 1141 GAAACAGAGCTACTCTCAAGAGGAAATCACTCAAGTTTGAAGAGAGAGGACACGCT 1200
DB GAAACAGAGCTACTCTCAAGAGGAAATCACTCAAGTTTGAAGAGAGAGGACACGCT 1200

QY 1201 CTATGCCCTTATCCACATTTTTCAGCTGCAGAACCAAGCTTCCCAAAAGAAATTTCTC 1260
DB CTATGCCCTTATCCACATTTTTCAGCTGCAGAACCAAGCTTCCCAAAAGAAATTTCTC 1260

QY 1261 CATCCACGGGCGCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAGAAAGAGG 1320
DB CATCCACGGGCGCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAGAAAGAGG 1320

QY 1321 CAGCTAACTCTCTCTGAAGAGTGGACAAATGTCAGCCGGCGGAGATCTAGCGGGAGC 1380
DB CAGCTAACTCTCTCTGAAGAGTGGACAAATGTCAGCCGGCGGAGATCTAGCGGGAGC 1380

QY 1381 TCAAAAGGATGTTGGCGAAATCTTGAGTCTTCTGAGAAAACCTGTACAAAGACTACGGGA 1440
DB TCAAAAGGATGTTGGCGAAATCTTGAGTCTTCTGAGAAAACCTGTACAAAGACTACGGGA 1440

QY 1441 ACAGTTTGGCTCCCTCCAGCCCTCAACCAATTTCTTCCATGCTGGGCTGATGGGCT 1500
DB ACAGTTTGGCTCCCTCCAGCCCTCAACCAATTTCTTCCATGCTGGGCTGATGGGCT 1500

QY 1501 AGTAAGACTTCCAGTTCTTTAGAGCGCTGTAGTATTTTCTTCTCTCATCTCTCTTGG 1560
DB AGTAAGACTTCCAGTTCTTTAGAGCGCTGTAGTATTTTCTTCTCTCATCTCTCTTGG 1560

QY 1561 ATACTTCTTTTAAAGTGGAGTCTCAGGCAACTCAAGTTTGTAGACCTTACTCTTTTGT 1620
DB ATACTTCTTTTAAAGTGGAGTCTCAGGCAACTCAAGTTTGTAGACCTTACTCTTTTGT 1620

QY 1621 GTTTTTTGAACAGGATCTTGTCTGTACCCAGGCTTGAAGTGGTGGATCACAG 1680
DB GTTTTTTGAACAGGATCTTGTCTGTACCCAGGCTTGAAGTGGTGGATCACAG 1680

QY 1681 CCAGTGCAGCTCGAACCACTGTGTCTCAAGCAATCTCCCATCTCCCATCTCCCAAAGTG 1740
DB CCAGTGCAGCTCGAACCACTGTGTCTCAAGCAATCTCCCATCTCCCATCTCCCAAAGTG 1740

QY 1741 CTGGGATGACAGGCTGAGCCACAGCTCCAGCTAGGCGCTTAATCTTCTGCTGTTATTTT 1800
DB CTGGGATGACAGGCTGAGCCACAGCTCCAGCTAGGCGCTTAATCTTCTGCTGTTATTTT 1800

QY 1801 CCATGAGCTAAAGGTCTGGTCTATCTGAGCTCACGCTGGCTCACACAGCTCTAGGGGCTG 1860





QY 721 CATGAATAGGACCAACGTCATGTCTTTTCTGAGCTTTCCGCTCTCTGTCGCAATGAAGA 780  
DB 721 CATGAATAGGACCAACGTCATGTCTTTTCTGAGCTTTCCGCTCTCTGTCGCAATGAAGA 780  
QY 781 CTTTCTGCTCCTGCTCACCCTAGCTCCTCTCTTGTGATGGCGCTGACCTTCTCATGTCCTC 840  
DB 781 CTTTGTCTCCTGCTCACCCTAGCTCCTCTCTTGTGATGGCGCTGACCTTCTCATGTCCTC 840  
QY 841 CTTTCACTTCTGCTGTTCTTTCACGGGCTGGAAGAGACATGCGGGCCCACTACTACTCAC 900  
DB 841 CTTTCACTTCTGCTGTTCTTTCACGGGCTGGAAGAGACATGCGGGCCCACTACTACTCAC 900  
QY 901 GATGCTCCTCTCATTTGCCATCTGGGTGCTGGATCACCTGCTCATGCTTCTTCTGACTTT 960  
DB 901 GATGCTCCTCTCATTTGCCATCTGGGTGCTGGATCACCTGCTCATGCTTCTTCTGACTTT 960  
QY 961 TGACCGCAGGTGGGATGACACCATCTCAGCTCCGCTTGGCTGGCTGCAATGGCTGGGTGTT 1020  
DB 961 TGACCGCAGGTGGGATGACACCATCTCAGCTCCGCTTGGCTGGCTGCAATGGCTGGGTGTT 1020  
QY 1021 CTTGTTGGCTTATGTTAGTCCGAGTTTGGCTGCTCACAAGCAACGAAACCCCATGGA 1080  
DB 1021 CTTGTTGGCTTATGTTAGTCCGAGTTTGGCTGCTCACAAGCAACGAAACCCCATGGA 1080  
QY 1081 TTATCCTGTTGAGGATGCTTTCTGTAACCTCAACTCGTGAAGAGAGCTATGTTGTGA 1140  
DB 1081 TTATCCTGTTGAGGATGCTTTCTGTAACCTCAACTCGTGAAGAGAGCTATGTTGTGA 1140  
QY 1141 GAAACAGAGCTACTCTCAAGAGGAAATCACTCAAGTTTTTGAAGAGACAGGGGACACGCT 1200  
DB 1141 GAAACAGAGCTACTCTCAAGAGGAAATCACTCAAGTTTTTGAAGAGACAGGGGACACGCT 1200  
QY 1201 CTATGCCCCCTATTCCACACATTTTCAGCTGAGAACACGACCTCCCAAAAGGAATTTCTC 1260  
DB 1201 CTATGCCCCCTATTCCACACATTTTCAGCTGAGAACACGACCTCCCAAAAGGAATTTCTC 1260  
QY 1261 CATCCACGGGCCACGCTTGGCCGAGCCCTTACAAAGCTATGAAGTAAAGAAAGAGGG 1320  
DB 1261 CATCCACGGGCCACGCTTGGCCGAGCCCTTACAAAGCTATGAAGTAAAGAAAGAGGG 1320  
QY 1321 CAGCTAACTCTCTCTGAAGAGTGGCAAAATGACGCGGGCGGAGATCTAGCGGGAGC 1380  
DB 1321 CAGCTAACTCTCTCTGAAGAGTGGCAAAATGACGCGGGCGGAGATCTAGCGGGAGC 1380  
QY 1381 TCAAAAGGATGTCGCGGAATCTTGAGTCTTCTGAGAAACTGTACAAGACACTACGGGA 1440  
DB 1381 TCAAAAGGATGTCGCGGAATCTTGAGTCTTCTGAGAAACTGTACAAGACACTACGGGA 1440  
QY 1441 ACAGTTTGCTCCCTCCAGCCTCAACCAATTTCTTCCATGCTGGGGCTGATGGGCT 1500  
DB 1441 ACAGTTTGCTCCCTCCAGCCTCAACCAATTTCTTCCATGCTGGGGCTGATGGGCT 1500  
QY 1501 AGTAAGACTCCAGTTCTTAGAGCGCTGAGTATTTTTTTTTTTTGTCTCATCTCTTTGG 1560  
DB 1501 AGTAAGACTCCAGTTCTTAGAGCGCTGAGTATTTTTTTTTTTTGTCTCATCTTTGG 1560  
QY 1561 ATACTCTTTTAAAGTGGAGTCTCAGGCAACTCAAGTTTGTAGCCCTTACTCTTTTGTGTT 1620  
DB 1561 ATACTCTTTTAAAGTGGAGTCTCAGGCAACTCAAGTTTGTAGCCCTTACTCTTTTGTGTT 1620  
QY 1621 GTTTTTTGAACAGGATCTTGCTCTGTCAACGAGCTTGAGTGCAGTGGTGCATCACAG 1680  
DB 1621 GTTTTTTGAACAGGATCTTGCTCTGTCAACGAGCTTGAGTGCAGTGGTGCATCACAG 1680  
QY 1681 CCCAGTGCAGCTTCACACACCTGTGCTCAAGCAATCTCCCATCTCTCCCAAGT 1740  
DB 1681 CCCAGTGCAGCTTCACACACCTGTGCTCAAGCAATCTCCCATCTCTCCCAAGT 1740  
QY 1741 CTGGGATGACAGGCGTGAAGCCACAGCTCCAGCCCTAGGCCCTTAATCTTGTGTTATTTT 1800  
DB 1741 CTGGGATGACAGGCGTGAAGCCACAGCTCCAGCCCTAGGCCCTTAATCTTGTGTTATTTT 1800  
QY 1801 CCATGACTAAAGTCTGTGTCATCTGAGCTCAGCTGCTCAGCTGCTCAGAGCTCTAGGGCCCTG 1860

DB 1801 CCATGACTAAAGTCTGTGTCATCTGAGCTCAGCTGCGCTCACAGCTCTTAGGGCCCTG 1860  
QY 1861 CTCCTCTAACTCACAGTGGGTTTTGTGAGGCTCTGTGGCCAGAGCAGACCTGTCATATCT 1920  
DB 1861 CTCCTCTAACTCACAGTGGGTTTTGTGAGGCTCTGTGGCCAGAGCAGACCTGTCATATCT 1920  
QY 1921 GAGCAAAAATAGCAAAAGCCTCTCTCAGCCCACTGGCCTGAAATCTACACTGGAAGCAAC 1980  
DB 1921 GAGCAAAAATAGCAAAAGCCTCTCTCAGCCCACTGGCCTGAAATCTACACTGGAAGCAAC 1980  
QY 1981 TTGCTGGCACCCTCCGCTCCCAACCTTCTTGGCTGGGTAGGAGGCTAAAGATCACCC 2040  
DB 1981 TTGCTGGCACCCTCCGCTCCCAACCTTCTTGGCTGGGTAGGAGGCTAAAGATCACCC 2040  
QY 2041 TAAATTTTACTCATCTCTCTAGTGTGCTCACAATTTGGGCTCAGCAGCTCCCGCAGCACA 2100  
DB 2041 TAAATTTTACTCATCTCTCTAGTGTGCTCACAATTTGGGCTCAGCAGCTCCCGCAGCACA 2100  
QY 2101 ATTACAGGTCAACCTCTCTTCTTGGCACTGTCGCCAAACTTGTCTCAATTTCCGAGATC 2160  
DB 2101 ATTACAGGTCAACCTCTCTTCTTGGCACTGTCGCCAAACTTGTCTCAATTTCCGAGATC 2160  
QY 2161 TAACTCTCCCTACGCTCTGCCAGGAATTTCTTTCAGACCTCACTAGCAAGCCGGTTG 2220  
DB 2161 TAACTCTCCCTACGCTCTGCCAGGAATTTCTTTCAGACCTCACTAGCAAGCCGGTTG 2220  
QY 2221 CTCCTTTGTGAGGAAATTTGTAGATCATTTCAATTTCTTCAAAATTTCTGGGCTGATCTTCT 2280  
DB 2221 CTCCTTTGTGAGGAAATTTGTAGATCATTTCAATTTCTTCAAAATTTCTGGGCTGATCTTCT 2280  
QY 2281 CTCATCTTGCAACCCCAACCTCTGTAATAGATTTACCGCATTTACGGCTGCATTTCTGTAA 2340  
DB 2281 CTCATCTTGCAACCCCAACCTCTGTAATAGATTTACCGCATTTACGGCTGCATTTCTGTAA 2340  
QY 2341 GTGGCAGTGTCTCTTAATGAGGAGTGTTCATTTGTAATAAGTTATTTCACTGAGTAT 2400  
DB 2341 GTGGCAGTGTCTCTTAATGAGGAGTGTTCATTTGTAATAAGTTATTTCACTGAGTAT 2400  
QY 2401 GCAATAAAGATGTGTGGCCTCTTTTCATGTGTGGCAGCAAAAAA 2456  
DB 2401 GCAATAAAGATGTGTGGCCTCTTTTCATGTGTGGCAGCAAAAAA 2456

RESULT 9  
US-11-169-041-32  
; Sequence 32, Application US/11169041  
; Publication No. US20060019284A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: IDENTIFICATION OF POLYNUCLEOTIDES FOR PREDICTING ACTIVITY OF  
; TITLE OF INVENTION: COMPOUNDS THAT INTERACT WITH AND/OR MODULATE PROTEIN TYROSINE  
; TITLE OF INVENTION: KINASES AND/OR PROTEIN TYROSINE KINASE PATHWAYS IN LUNG CANCER  
; TITLE OF INVENTION: CELLS  
; FILE REFERENCE: 10001 NP  
; CURRENT APPLICATION NUMBER: US/11/169,041  
; CURRENT FILING DATE: 2005-06-28  
; PRIOR APPLICATION NUMBER: 60/584,405  
; PRIOR FILING DATE: 2004-06-30  
; NUMBER OF SEQ ID NOS: 527  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 32  
; LENGTH: 2456  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-169-041-32

Query Match 100.0%; Score 2456; DB 16; Length 2456;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATAAAGCATGAAGTGGCGTGAACCTGGAATAGGCGTGTCTCTCCCTCGACCCCTCCCC 60

Db 1 ATAACAGCATGAAGTCGCGTGAACCTGGAATAGGCGGTCTCTCTCCCTCGACCCCTCCCC 60  
Qy 61 TCCTGTCCCTCTGCTCACCCTCGCTCGTTCCTCCCTCGGGAGGCGCGCTTTTATA 120  
Db 61 TCCTGTCCCTCTGCTCACCCTCGCTCGTTCCTCCCTCGGGAGGCGCGCTTTTATA 120  
Qy 121 ACAAATGCTCAGAGTCGAGGGGGGATAGCTGTCCAAAGGTCTCCCCAGCACTGAGGAG 180  
Db 121 ACAAATGCTCAGAGTCGAGGGGGGATAGCTGTCCAAAGGTCTCCCCAGCACTGAGGAG 180  
Qy 181 CTGCGCTGTGCCCTCTTTGCGCGGGGAAGACAGACAAAGTTCA CGGCCAAACCCCTTGGC 240  
Db 181 CTGCGCTGTGCCCTCTTTGCGCGGGGAAGACAGACAAAGTTCA CGGCCAAACCCCTTGGC 240  
Qy 241 ACTAGGGTCAGAAATGCTACACAGCTCCCTGATGTTGCGCAATGGCCGTGAATCCAA 300  
Db 241 ACTAGGGTCAGAAATGCTACACAGCTCCCTGATGTTGCGCAATGGCCGTGAATCCAA 300  
Qy 301 GTACTACAGACTTTGTGATAAGGCTGAAGCTTGGGGCATCGTCTAGAAAACGGTGGCCAC 360  
Db 301 GTACTACAGACTTTGTGATAAGGCTGAAGCTTGGGGCATCGTCTAGAAAACGGTGGCCAC 360  
Qy 361 AGCGGGGTTGTGACCTCGGTGCGCTTCACTCTCCCGATCCTCGTCTGCAAGGT 420  
Db 361 AGCGGGGTTGTGACCTCGGTGCGCTTCACTCTCCCGATCCTCGTCTGCAAGGT 420  
Qy 421 GCAGGACTCAACAGGGGAAAATGCTGCTACTCAGTTTCTCTCTCTCTGCGGTGTGT 480  
Db 421 GCAGGACTCAACAGGGGAAAATGCTGCTACTCAGTTTCTCTCTCTCTGCGGTGTGT 480  
Qy 481 GGGCATCTTTGGGCTCACTCTCGCTTCACTCGGACTGGAGGAGACAGGGCCAC 540  
Db 481 GGGCATCTTTGGGCTCACTCTCGCTTCACTCGGACTGGAGGAGACAGGGCCAC 540  
Qy 541 ACCTTTCTTCTCTTTGGGATCTCTTTTCCATCTGCTCTCTCGCTCGCTGCTCATGC 600  
Db 541 ACCTTTCTTCTCTTTGGGATCTCTTTTCCATCTGCTCTCTCGCTCGCTGCTCATGC 600  
Qy 601 TGTCACTGACCAAGCTCGTCCGGGGAGGAGCCCTTTCCCTGTTGTTGATCTTGGG 660  
Db 601 TGTCACTGACCAAGCTCGTCCGGGGAGGAGCCCTTTCCCTGTTGTTGATCTTGGG 660  
Qy 661 TCTGGCGGTGGGCTTCAAGCTAGTCCAGAGTGTATCGCTATTGAATATATATGCTCTGAC 720  
Db 661 TCTGGCGGTGGGCTTCAAGCTAGTCCAGAGTGTATCGCTATTGAATATATATGCTCTGAC 720  
Qy 721 CATGAATAGGACCAACGTCATATGCTTTCTGAGCTTTCGCTCCTCGTGCATGAAGA 780  
Db 721 CATGAATAGGACCAACGTCATATGCTTTCTGAGCTTTCGCTCCTCGTGCATGAAGA 780  
Qy 781 CTTTGTCTCTGCTCACCCTACGTCCTCTTCTTGTATGGCGCTGACCTTCTCATGTCTTC 840  
Db 781 CTTTGTCTCTGCTCACCCTACGTCCTCTTCTTGTATGGCGCTGACCTTCTCATGTCTTC 840  
Qy 841 CTTTCACTTCTGTGGTTCCTTCAAGGCTGGAAGAGACATGGGGCCCAATCTACCTCAC 900  
Db 841 CTTTCACTTCTGTGGTTCCTTCAAGGCTGGAAGAGACATGGGGCCCAATCTACCTCAC 900  
Qy 901 GATGCTCCTCTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960  
Db 901 GATGCTCCTCTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960  
Qy 961 TGACCGAGGTGGGATGACACCATCTCAGCTCGGCTTGGCTGGCCATGGGCTGGGTGT 1020  
Db 961 TGACCGAGGTGGGATGACACCATCTCAGCTCGGCTTGGCTGGCCATGGGCTGGGTGT 1020  
Qy 1021 CCGTGTGGCTTATGTTAGTCCCGAGTTTGGCTGCTCACAAGAACGAAACCCCATGGA 1080  
Db 1021 CCGTGTGGCTTATGTTAGTCCCGAGTTTGGCTGCTCACAAGAACGAAACCCCATGGA 1080  
Qy 1081 TTATCCTGTTGAGGATGCTTTCTGTAACCTCAACTCGTGAAGAGAGCTATGTTGTGA 1140  
Db 1081 TTATCCTGTTGAGGATGCTTTCTGTAACCTCAACTCGTGAAGAGAGCTATGTTGTGA 1140

Qy 1141 GAACAGAGCTACTCTCAAGAGGAAATCACTCAAGCTTTTGAAGACAGAGGACACGCT 1200  
Db 1141 GAACAGAGCTACTCTCAAGAGGAAATCACTCAAGCTTTTGAAGACAGAGGACACGCT 1200  
Qy 1201 CTATGCCCCCTATTTCACACATTTTTCAGCTGAGAACCCAGCTTCCCAAGAAATTTCTC 1260  
Db 1201 CTATGCCCCCTATTTCACACATTTTTCAGCTGAGAACCCAGCTTCCCAAGAAATTTCTC 1260  
Qy 1261 CATCCCAAGGGCCCAAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAGAAAGGG 1320  
Db 1261 CATCCCAAGGGCCCAAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAGAAAGGG 1320  
Qy 1321 CAGCTAACTCTGTCTGAAGAGTGGGACAAAATGACGCGGGCGGAGATCTACGGGAGC 1380  
Db 1321 CAGCTAACTCTGTCTGAAGAGTGGGACAAAATGACGCGGGCGGAGATCTACGGGAGC 1380  
Qy 1381 TCAAGGGGATGTGGGCGAAATCTTGAGTCTTCTGAGAAAATCTGTACAAGACACTACGGGA 1440  
Db 1381 TCAAGGGGATGTGGGCGAAATCTTGAGTCTTCTGAGAAAATCTGTACAAGACACTACGGGA 1440  
Qy 1441 ACAGTTTGGCTTCCCTCCAGCCCTCAACCAATTTCTTCATGCTGGGGCTGATGGGCT 1500  
Db 1441 ACAGTTTGGCTTCCCTCCAGCCCTCAACCAATTTCTTCATGCTGGGGCTGATGGGCT 1500  
Qy 1501 AGTAAGACTCCAGTCTTAGAGCGCTGTAGTATTTTTTTTTTTTGTCTCATCTTTGG 1560  
Db 1501 AGTAAGACTCCAGTCTTAGAGCGCTGTAGTATTTTTTTTTTTTGTCTCATCTTTGG 1560  
Qy 1561 ATACTTCTTTTAAAGTGGGAGTCTCAGGCAACTCAAGTTTAGACCTTACTCTTTTGT 1620  
Db 1561 ATACTTCTTTTAAAGTGGGAGTCTCAGGCAACTCAAGTTTAGACCTTACTCTTTTGT 1620  
Qy 1621 GTTTTTTGAACAGGATCTTGTCTGTCAACCAGGCTTGAGTGCAGTGTGCGATCACAG 1680  
Db 1621 GTTTTTTGAACAGGATCTTGTCTGTCAACCAGGCTTGAGTGCAGTGTGCGATCACAG 1680  
Qy 1681 CCAGTGCAGCTCCAGCACCTGTCTCAAGCAATCTCCCATCTCCATCTCCCAAGTG 1740  
Db 1681 CCAGTGCAGCTCCAGCACCTGTCTCAAGCAATCTCCCATCTCCATCTCCCAAGTG 1740  
Qy 1741 CTGGGATGACAGCGCTGAGCCACAGCTCCAGCTAGGCGCTTAAATCTTGTCTTATTT 1800  
Db 1741 CTGGGATGACAGCGCTGAGCCACAGCTCCAGCTAGGCGCTTAAATCTTGTCTTATTT 1800  
Qy 1801 CATGACTTAAAGGTCTGTCTGAGCTCAGCTGCGTCAACAGCTCTAGGGGCTG 1860  
Db 1801 CATGACTTAAAGGTCTGTCTGAGCTCAGCTGCGTCAACAGCTCTAGGGGCTG 1860  
Qy 1861 CTCCTTAACTCAGTGGGTTTGTGAGGCTCTGTGGCCAGAGCAGACCTGCATATCT 1920  
Db 1861 CTCCTTAACTCAGTGGGTTTGTGAGGCTCTGTGGCCAGAGCAGACCTGCATATCT 1920  
Qy 1921 GAGCAAAAATAGCAAAAGCTCTCTCAGCCCACTGGCCTGAATCTACACTGGAAGCAAC 1980  
Db 1921 GAGCAAAAATAGCAAAAGCTCTCTCAGCCCACTGGCCTGAATCTACACTGGAAGCAAC 1980  
Qy 1981 TTGCTGGCAACCCCGCTCCCAACCCCTTCTGCTGGGTAGGAGGCTTAAAGATCACCC 2040  
Db 1981 TTGCTGGCAACCCCGCTCCCAACCCCTTCTGCTGGGTAGGAGGCTTAAAGATCACCC 2040  
Qy 2041 TAAATTTACTCATCTCTAGTGTGCTCACAATGGGCTCAGAGCTCCCGAGCACA 2100  
Db 2041 TAAATTTACTCATCTCTAGTGTGCTCACAATGGGCTCAGAGCTCCCGAGCACA 2100  
Qy 2101 ATTACAGGTCAACCCCTCTCTTGTGCACTGTGCCAAAATCTTGTCTCAATTCGAGATC 2160  
Db 2101 ATTACAGGTCAACCCCTCTCTTGTGCACTGTGCCAAAATCTTGTCTCAATTCGAGATC 2160  
Qy 2161 TAAATCTCCCTTACGCTCTGCGAGGAAATCTTTTCAAGCTCTACTAGCAAGCCCGGTTG 2220  
Db 2161 TAAATCTCCCTTACGCTCTGCGAGGAAATCTTTTCAAGCTCTACTAGCAAGCCCGGTTG 2220

QY 2221 CTCCTGTGTCAGAGAAATTTGTAGATCATTTCTCACTTCAAATTCCTGGGCTGATACTTCT 2280  
|||  
Db 2221 CTCCTGTGTCAGAGAAATTTGTAGATCATTTCTCACTTCAAATTCCTGGGCTGATACTTCT 2280  
|||  
QY 2281 CTCATCTTCACCCCAACCTCTGTAAATAGATTTACCGCATTTACGGCTGCAATTCGTAA 2340  
|||  
Db 2281 CTCATCTTCACCCCAACCTCTGTAAATAGATTTACCGCATTTACGGCTGCAATTCGTAA 2340  
|||  
QY 2341 GTGGCATGGTCTCCTAAATGGAGGAGTGTTCATTTGATATAAAGTTATTCACTCGATAT 2400  
|||  
Db 2341 GTGGCATGGTCTCCTAAATGGAGGAGTGTTCATTTGATATAAAGTTATTCACTCGATAT 2400  
|||  
QY 2401 GCAATAAAGATGTGGTGGCCACTCTTTTCATGGTGGGAGCAAAAAA 2456  
|||  
Db 2401 GCAATAAAGATGTGGTGGCCACTCTTTTCATGGTGGGAGCAAAAAA 2456  
|||

RESULT 10

US-10-600-816-18  
; Sequence 18, Application US/10600816  
; Publication No. US20040121362A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: IDENTIFICATION AND MODULATION OF A G-PROTEIN COUPLED RECEPTOR  
; TITLE OF INVENTION: (GPCR), RAI3, ASSOCIATED WITH CHRONIC OBSTRUCTIVE PULMONARY  
; TITLE OF INVENTION: DISEASE (COPD) AND NF-KB AND E-SELECTIN REGULATION  
; FILE REFERENCE: D0251 NP  
; CURRENT APPLICATION NUMBER: US/10/600,816  
; PRIOR FILING DATE: 2003-06-20  
; PRIOR APPLICATION NUMBER: U.S. 60/390,850  
; PRIOR FILING DATE: 2002-06-20  
; PRIOR APPLICATION NUMBER: U.S. 60/407,006  
; PRIOR FILING DATE: 2002-08-29  
; NUMBER OF SEQ ID NOS: 98  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 18  
; LENGTH: 2456  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: RAI3 Polymorphic Allele Summary Sequence.  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (112)..(112)  
; OTHER INFORMATION: wherein "n" equals either G or A.  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (364)..(364)  
; OTHER INFORMATION: wherein "n" equals either C or T.  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (511)..(511)  
; OTHER INFORMATION: wherein "n" equals either C or T.  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (523)..(523)  
; OTHER INFORMATION: wherein "n" equals either C or T.  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (605)..(605)  
; OTHER INFORMATION: wherein "n" equals either A or G.  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (797)..(797)  
; OTHER INFORMATION: wherein "n" equals either A or G.  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (111)..(111)  
; OTHER INFORMATION: wherein "n" equals either T or C.  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1173)..(1173)  
; OTHER INFORMATION: wherein "n" equals either A or G.

US-10-600-816-18  
Query Match 99.7%; Score 2448; DB 8; Length 2456;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 2448; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
QY 1 ATAAAGCATGAGTGCCTGGAACTGGAAATAGGCGTGTCTCTCCCTCGAGCCCTCCCC 60  
Db 1 ATAAAGCATGAGTGCCTGGAACTGGAAATAGGCGTGTCTCTCCCTCGAGCCCTCCCC 60  
QY 61 TCCTTGTCCCTCTGCTCACCCCTCGCTCCCTCCCTCCCGGAGGCGCGCTTTATA 120  
Db 61 TCCTTGTCCCTCTGCTCACCCCTCGCTCCCTCCCTCCCGGAGGCGCGCTTTATA 120  
QY 121 ACAACTGCTCAGAGTGCAGAGGCGGGAATAGTGTCAAAGGTCTCCCGCAGCACTGAGAG 180  
Db 121 ACAACTGCTCAGAGTGCAGAGGCGGGAATAGTGTCAAAGGTCTCCCGCAGCACTGAGAG 180  
QY 181 CTCGCTGCTGCCCTCTTTCGCGCGGGAAGCAGCACCAGTTTCAGGGCAACGCCCTTGGC 240  
Db 181 CTCGCTGCTGCCCTCTTTCGCGCGGGAAGCAGCACCAGTTTCAGGGCAACGCCCTTGGC 240  
QY 241 ACTAGGGTCCAGATGGCTACAACAGTCCCTGATGGTTGCCGCAATGGCTGAAATCCAA 300  
Db 241 ACTAGGGTCCAGATGGCTACAACAGTCCCTGATGGTTGCCGCAATGGCTGAAATCCAA 300  
QY 301 GTACTACAGACTTTGTGATAAGGCTGAAAGCTTTGGGGCATCTGCTCTAGAAAACGGTGGCCAC 360  
Db 301 GTACTACAGACTTTGTGATAAGGCTGAAAGCTTTGGGGCATCTGCTCTAGAAAACGGTGGCCAC 360  
QY 361 AGCGGGGTGTGACCTCGGTGGCTCATGCTCACTCTCCCGATCCCTCGTGAAGGT 420  
Db 361 AGCGGGGTGTGACCTCGGTGGCTCATGCTCACTCTCCCGATCCCTCGTGAAGGT 420  
QY 421 GCAGGACTCCAAACAGGCGAAAAATGCTGCTACTCAGATTTCTCTCTCTCTGGGTGTGT 480  
Db 421 GCAGGACTCCAAACAGGCGAAAAATGCTGCTACTCAGATTTCTCTCTCTCTGGGTGTGT 480  
QY 481 GGGCATCTTTGGCTTCACCTTCGCTTCATCATCGGACTGGAGCGGAGCAGAGGCCAC 540  
Db 481 GGGCATCTTTGGCTTCACCTTCGCTTCATCATCGGACTGGAGCGGAGCAGAGGCCAC 540  
QY 541 AGGCTTCTTCTCTTTGGATCCTCTTTTCCATCTGCTTCTCTGCTGCTGCTCATGC 600  
Db 541 AGGCTTCTTCTCTTTGGATCCTCTTTTCCATCTGCTTCTCTGCTGCTGCTCATGC 600  
QY 601 TGTCACTGTGACCAAGCTCGTCCGGGGAGGAAGCCCTTTCCCTGTGGTGTATCTGGG 660  
Db 601 TGTCACTGTGACCAAGCTCGTCCGGGGAGGAAGCCCTTTCCCTGTGGTGTATCTGGG 660  
QY 661 TGTGCGGTGGCTTTCAGCTTACAGATGTTATCGCTATTTGAATATATATTTCTGTAC 720  
Db 661 TGTGCGGTGGCTTTCAGCTTACAGATGTTATCGCTATTTGAATATATATTTCTGTAC 720  
QY 721 CATGAATAGGACCAACGTCATGCTTTTCTCAGCTTTTCCGCTCTCGTCAATGAAGA 780  
Db 721 CATGAATAGGACCAACGTCATGCTTTTCTCAGCTTTTCCGCTCTCGTCAATGAAGA 780  
QY 781 CTTTGTCTCTGCTCACCTACGCTCTCTCTTGTGATGGCGTGAACCTTCTCATGTCTTC 840  
Db 781 CTTTGTCTCTGCTCNCCTAGCTCTCTCTTGTGATGGCGTGAACCTTCTCATGTCTTC 840  
QY 841 CTTCACTTCTGTGTTTCTTTCAGGGTGGAGAGACATGGGGGCCACATCTACCTCAC 900  
Db 841 CTTCACTTCTGTGTTTCTTTCAGGGTGGAGAGACATGGGGGCCACATCTACCTCAC 900  
QY 901 GATGCTCTCTCCATTGCCATCTGCTGCTGGATCACCTGCTCATCTCTCTCTGACTT 960  
Db 901 GATGCTCTCTCCATTGCCATCTGCTGCTGGATCACCTGCTCATCTCTCTCTGACTT 960  
QY 961 TGACCGCAGGTGGGATGACACCATCTCAGCTCCGCTTGGCTGCGCAATGGGTGGTGT 1020  
Db 961 TGACCGCAGGTGGGATGACACCATCTCAGCTCCGCTTGGCTGCGCAATGGGTGGTGT 1020

```
QY 1021 CTTGTTGGCTTATGTTAGTCCCGAGTTTGGCTGCTCAGAAAACGAAACCCCATGGA 1080
Db 1021 CTTGTTGGCTTATGTTAGTCCCGAGTTTGGCTGCTCAGAAAACGAAACCCCATGGA 1080
QY 1081 TTATCTGTTGAGGATGCTTTCTGTAACCTCAACTCGTGAAGAGAGCTATGCTGTGGA 1140
Db 1081 TTATCTGTTGAGGATGCTTTCTGTAACCTCAACTCGTGAAGAGAGCTATGCTGTGGA 1140
QY 1141 GAACAGAGCTACTCTCAAGAGAAATCACTCAAGGTTTGAAGAGACAGGGACAGCT 1200
Db 1141 GAACAGAGCTACTCTCAAGAGAAATCACTCAAGGTTTGAAGAGACAGGGACAGCT 1200
QY 1201 CTATGCCCCCTATTCCACACATTTTCAGCTGCAGAACCCAGCTCCCAAGAAAGAAATTC 1260
Db 1201 CTATGCCCCCTATTCCACACATTTTCAGCTGCAGAACCCAGCTCCCAAGAAAGAAATTC 1260
QY 1261 CATCCACAGGGCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAGAAAGAGGG 1320
Db 1261 CATCCACAGGGCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAGAAAGAGGG 1320
QY 1321 CAGCTAACTCTGCTGAAGAGTGGACAAAATGCAGCCGGCGGCGAGATCTAGCGGGAGC 1380
Db 1321 CAGCTAACTCTGCTGAAGAGTGGACAAAATGCAGCCGGCGGCGAGATCTAGCGGGAGC 1380
QY 1381 TCAAGGGGATGGGGCGAAATCTTGAGTCTTCTGAGAAAACCTGTACAAGACACTACGGGA 1440
Db 1381 TCAAGGGGATGGGGCGAAATCTTGAGTCTTCTGAGAAAACCTGTACAAGACACTACGGGA 1440
QY 1441 ACAGTTTGCTCCCTCCCGAGCTCAACCAATTTCTTCCATGCTGGGGCTGATGGGCT 1500
Db 1441 ACAGTTTGCTCCCTCCCGAGCTCAACCAATTTCTTCCATGCTGGGGCTGATGGGCT 1500
QY 1501 AGTAAGACTCCAGTCTTAGAGCGCTGATGATATTTTTTTTTTTTGTCTCATCTTTGG 1560
Db 1501 AGTAAGACTCCAGTCTTAGAGCGCTGATGATATTTTTTTTTTTTGTCTCATCTTTGG 1560
QY 1561 ATACTTCTTTTAAAGTGGAGTCTCAGGCAACTCAAGTTTAGACCCCTACTCTTTTGGTT 1620
Db 1561 ATACTTCTTTTAAAGTGGAGTCTCAGGCAACTCAAGTTTAGACCCCTACTCTTTTGGTT 1620
QY 1621 GTTTTTTGAACAGAGTCTTTGCTCTGTCAACCGAGCTTGAAGTGGTGGATCACAG 1680
Db 1621 GTTTTTTGAACAGAGTCTTTGCTCTGTCAACCGAGCTTGAAGTGGTGGATCACAG 1680
QY 1681 CCCAGTGCAGCTCGACACCTGTGCTCAAGCAATCTCCCATCTCCATCTCCCAAGTG 1740
Db 1681 CCCAGTGCAGCTCGACACCTGTGCTCAAGCAATCTCCCATCTCCATCTCCCAAGTG 1740
QY 1741 CTGGGATGACGGCGTGAAGCTCCAGCTCCAGCTAGGCGCTTAATCTTGTGTTATTTT 1800
Db 1741 CTGGGATGACGGCGTGAAGCTCCAGCTAGGCGCTTAATCTTGTGTTATTTT 1800
QY 1801 CCATGACTAAAGGCTGTGCTCATCTCAGCTCAGCTGGCTCACACAGCTCTAGGGCCCTG 1860
Db 1801 CCATGACTAAAGGCTGTGCTCATCTCAGCTCAGCTGGCTCACACAGCTCTAGGGCCCTG 1860
QY 1861 CTCCTCTAATCAGTGGGTTTGTGAGGCTCTGTGGCCGAGAGCAGACCTGCTATATCT 1920
Db 1861 CTCCTCTAATCAGTGGGTTTGTGAGGCTCTGTGGCCGAGAGCAGACCTGCTATATCT 1920
QY 1921 GAGCAAAATAGCAAAAGCTCTCTCAGCCCACTGGCCCTGAATCTACACTGGAAGCCAC 1980
Db 1921 GAGCAAAATAGCAAAAGCTCTCTCAGCCCACTGGCCCTGAATCTACACTGGAAGCCAC 1980
QY 1981 TTGCTGGCACCCCGCTCCCAACCTTCTTGCTGGTAGGAGAGGCTAAAGATCACCC 2040
Db 1981 TTGCTGGCACCCCGCTCCCAACCTTCTTGCTGGTAGGAGAGGCTAAAGATCACCC 2040
QY 2041 TAAATTTACTCATCTCTAGTGTGCTCACAATTTGGGCTCAGAGCTCCCGCAGCACCA 2100
Db 2041 TAAATTTACTCATCTCTAGTGTGCTCACAATTTGGGCTCAGAGCTCCCGCAGCACCA 2100
```

```
QY 2101 ATTCAGAGTCAACCCCTCTCTTCTGCACTGTGCCCAAACTTGTCTGTCATTCGAGATC 2160
Db 2101 ATTCAGAGTCAACCCCTCTCTTCTGCACTGTGCCCAAACTTGTCTGTCATTCGAGATC 2160
QY 2161 TAATCTCCCTCACTACGCTCTGCCAGGAATTTCTTTAGACCTCTACTAGCACAGCCCGTTG 2220
Db 2161 TAATCTCCCTCACTACGCTCTGCCAGGAATTTCTTTAGACCTCTACTAGCACAGCCCGTTG 2220
QY 2221 CTCCTTGTGAGAGAAATTTGTAGATCAATTCACATTTCAAAATTCCTGGGCTGATCTCT 2280
Db 2221 CTCCTTGTGAGAGAAATTTGTAGATCAATTCACATTTCAAAATTCCTGGGCTGATCTCT 2280
QY 2281 CTCATCTTGACCCCAACCTCTGTAATAGATTTACCGCATTTACGGCTGCAATTCCTGTAA 2340
Db 2281 CTCATCTTGACCCCAACCTCTGTAATAGATTTACCGCATTTACGGCTGCAATTCCTGTAA 2340
QY 2341 GTGGCATGCTCTCTTAATGGAGAGTGTTCATTTGATTAATAGTTTATTCACCTGAGTAT 2400
Db 2341 GTGGCATGCTCTCTTAATGGAGAGTGTTCATTTGATTAATAGTTTATTCACCTGAGTAT 2400
QY 2401 GCAATAAAGATGTGGTGGCCACTCTTTTCATGTTGGTGGCAGCAAAAAA 2456
Db 2401 GCAATAAAGATGTGGTGGCCACTCTTTTCATGTTGGTGGCAGCAAAAAA 2456

RESULT 11
US-10-198-846-10424
; Sequence 10424, Application US/10198846
; Publication No. US2003099974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10424
; LENGTH: 4239
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1, 4238, 4239
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-10424

Query Match 99.5%; Score 2443; DB 6; Length 4239;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2443; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATACAGCATGAAGTCCGTGGAACCTGGAATAGGCGTGTCTCTCCCTCGACCCCTCCCC 60
Db 409 ATACAGCATGAAGTCCGTGGAACCTGGAATAGGCGTGTCTCTCCCTCGACCCCTCCCC 468
QY 61 TCCTTGTCCCTGTCTCACCCTCGCTCGTTTCCCTCCCTCCGGCAGGGCCGCTTTATA 120
Db 469 TCCTTGTCCCTGTCTCACCCTCGCTCGTTTCCCTCCCTCCGGCAGGGCCGCTTTATA 528
QY 121 ACAATGCTCAGATGCGAGGGCGGATAGCTGTCCAAGTCTCCCGCCAGCACTGAGGAG 180
Db 529 ACAATGCTCAGATGCGAGGGCGGATAGCTGTCCAAGTCTCCCGCCAGCACTGAGGAG 588
QY 181 CTCGCTGTGCTTCTTTCGCGCGGGAAGCAGCAACAAAGTTTACGGCCAAAGCCTTGGC 240
Db 589 CTCGCTGTGCTTCTTTCGCGCGGGAAGCAGCAACAAAGTTTACGGCCAAAGCCTTGGC 648
```

QY	241	ACTAGGGTCCAGATGGCTACAACAGTCCCTCATGAGTTGCCCAATGGGCTGAATAATCCAA	300
Db	649	ACTAGGGTCCAGATGGCTACAACAGTCCCTCATGAGTTGCCCAATGGGCTGAATAATCCAA	708
QY	301	GTACTACAGACTTTGTGATAAGGCTGAAGCTTGGGGCATCGCTTAGAAGCGTGGCCAC	360
Db	709	GTACTACAGACTTTGTGATAAGGCTGAAGCTTGGGGCATCGCTTAGAAGCGTGGCCAC	768
QY	361	AGCCGGGTTGTGACCTCCGTTGGCCTTTCATGCTCACTCTCCGATCCTCGTCTGCAAGGT	420
Db	769	AGCCGGGTTGTGACCTCCGTTGGCCTTTCATGCTCACTCTCCGATCCTCGTCTGCAAGGT	828
QY	421	GCAGGACTCCAAAGCGGAAAAATGCTGCTACTCAGTTTCTCTTCCTCCCTGGGTGTGT	480
Db	829	GCAGGACTCCAAAGCGGAAAAATGCTGCTACTCAGTTTCTCTTCCTCCCTGGGTGTGT	888
QY	481	GGGCATCTTTGGCCTCACCTTGGCCTTTCATCATCGGACTGGAGCGGAGCACAGGGCCAC	540
Db	889	GGGCATCTTTGGCCTCACCTTGGCCTTTCATCATCGGACTGGAGCGGAGCACAGGGCCAC	948
QY	541	AGCCTTCTTCCTTTGGGATCCTCTTTTCCATCTGCTTCTCTGCTGCTGGCTCATGC	600
Db	949	AGCCTTCTTCCTTTGGGATCCTCTTTTCCATCTGCTTCTCTGCTGCTGGCTCATGC	1008
QY	601	TGTCAGTGTGACCAAGCTCGTCCGGGGAGGAAGCCCTTTCCCTGTGGTGAATCTGGG	660
Db	1009	TGTCAGTGTGACCAAGCTCGTCCGGGGAGGAAGCCCTTTCCCTGTGGTGAATCTGGG	1068
QY	661	TCGCGCGTGGGCTTTCAGCCTAGTCCAGGATGTTATCGCTATATGATATATTTGCTCTGAC	720
Db	1069	TCGCGCGTGGGCTTTCAGCCTAGTCCAGGATGTTATCGCTATATGATATATTTGCTCTGAC	1128
QY	721	CATGAATAGGACCAAGCTCAATGCTTTTCTCAGCTTTCCGCTCTCGTCGCAATGAAGA	780
Db	1129	CATGAATAGGACCAAGCTCAATGCTTTTCTCAGCTTTCCGCTCTCGTCGCAATGAAGA	1188
QY	781	CTTTGCTCTCTGCTCACCTACCTCTCTCTTTGATGGCGCTGACCTTCTCATGTCCTC	840
Db	1189	CTTTGCTCTCTGCTCACCTACCTCTCTCTTTGATGGCGCTGACCTTCTCATGTCCTC	1248
QY	841	CTTCACTTCTGTGGTTCCTTTCACGGCTGGAGACATGGGGCCACATCTACTCAC	900
Db	1249	CTTCACTTCTGTGGTTCCTTTCACGGCTGGAGACATGGGGCCACATCTACTCAC	1308
QY	901	GATGCTCTCTCCATTTGCCATCTGGGTGGCCTGGATCACCTGCTCATGCTTCTTGACTT	960
Db	1309	GATGCTCTCTCCATTTGCCATCTGGGTGGCCTGGATCACCTGCTCATGCTTCTTGACTT	1368
QY	961	TGACCGAGTGGGATGACACCATCTCAGCTCCGCTTTGGTGGCAATGGCTGGGTGTT	1020
Db	1369	TGACCGAGTGGGATGACACCATCTCAGCTCCGCTTTGGTGGCAATGGCTGGGTGTT	1428
QY	1021	CCTGTGGCTTATGTAGTCCGAGTTTGGCTGCTCACAAGACACGAACCCCATGGA	1080
Db	1429	CCTGTGGCTTATGTAGTCCGAGTTTGGCTGCTCACAAGACACGAACCCCATGGA	1488
QY	1081	TTATCTGTTGAGGATGCTTTTCTGTAACCTCAACTCGTGAAGAGAGATATGGTGTGGA	1140
Db	1489	TTATCTGTTGAGGATGCTTTTCTGTAACCTCAACTCGTGAAGAGAGATATGGTGTGGA	1548
QY	1141	GAACAGAGCTACTCTCAAGAGAAATCACTCAAGGTTTTGAAGAGACAGGGGACACGCT	1200
Db	1549	GAACAGAGCTACTCTCAAGAGAAATCACTCAAGGTTTTGAAGAGACAGGGGACACGCT	1608
QY	1201	CTATGCCCTTATTCACACATTTTCAGCTGAGACACGCTCCGCCAAAGAAATTTCTC	1260
Db	1609	CTATGCCCTTATTCACACATTTTCAGCTGAGAACACGCTCCGCCAAAGAAATTTCTC	1668
QY	1261	CATCCACCGGGCCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAGAAAGAGGG	1320
Db	1669	CATCCACCGGGCCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAGAAAGAGGG	1728

QY	1321	CAGCTAACTCTGCTGAAGAGTGGGACAAATGCGAGCCGGCGGCGAGATCTAGCGGGAGC	1380
Db	1729	CAGCTAACTCTGCTGAAGAGTGGGACAAATGCGAGCCGGCGGCGAGATCTAGCGGGAGC	1788
QY	1381	TCAAAAGGAGTGGCGGAAATCTTGAGTCTTCTGAGAAAACCTGTACAACACTACGGGA	1440
Db	1789	TCAAAAGGAGTGGCGGAAATCTTGAGTCTTCTGAGAAAACCTGTACAACACTACGGGA	1848
QY	1441	ACAGTTTGGCTCCCTCCAGCCCTCAACCAATTTCTTCCATGCTGGGGCTGATGTGGGCT	1500
Db	1849	ACAGTTTGGCTCCCTCCAGCCCTCAACCAATTTCTTCCATGCTGGGGCTGATGTGGGCT	1908
QY	1501	AGTAAGACTCAGTTCCTTAGAGGCGCTGATATTTTTTTTTTTTTTGTCTCATCTTTGG	1560
Db	1909	AGTAAGACTCAGTTCCTTAGAGGCGCTGATATTTTTTTTTTTTTTGTCTCATCTTTGG	1968
QY	1561	ATACTTCTTTTAAAGTGGGAGTCTCAGGCAACTCAAGTTTGTAGACCTTACTCTTTTGT	1620
Db	1969	ATACTTCTTTTAAAGTGGGAGTCTCAGGCAACTCAAGTTTGTAGACCTTACTCTTTTGT	2028
QY	1621	GTTTTTGAAACAGGATCTTGCTCTGTCAACCAGGCTTGAGTGCAGTGGTGCATCACAG	1680
Db	2029	GTTTTTGAAACAGGATCTTGCTCTGTCAACCAGGCTTGAGTGCAGTGGTGCATCACAG	2088
QY	1681	CCAGTGCAGCCTCGACACCTGTGCTCAAGCAATCTCTCCATCTCCATCTCCCAAGTG	1740
Db	2089	CCAGTGCAGCCTCGACACCTGTGCTCAAGCAATCTCTCCATCTCCATCTCCCAAGTG	2148
QY	1741	CTGGGATGACAGGCTGAGCCACAGCTCCAGCCTAGGCCCTTAATCTTGTGTTATTTT	1800
Db	2149	CTGGGATGACAGGCTGAGCCACAGCTCCAGCCTAGGCCCTTAATCTTGTGTTATTTT	2208
QY	1801	CCATGAGCTAAAGGCTCGTGTCTATCTGAGCTACGCTGCTCACACAGCTCTTAGGGCCTG	1860
Db	2209	CCATGAGCTAAAGGCTCGTGTCTATCTGAGCTACGCTGCTCACACAGCTCTTAGGGCCTG	2268
QY	1861	CTCCTCTAATCACAAGTGGGTTTTGTGAGGCTCTGTGGCCAGAGCAGCTGCATATCT	1920
Db	2269	CTCCTCTAATCACAAGTGGGTTTTGTGAGGCTCTGTGGCCAGAGCAGCTGCATATCT	2328
QY	1921	GAGCAAAATAGCAAAAGCCTCTCTCAGCCACTGGCCTGAATCTACACTGGAAGCAAC	1980
Db	2329	GAGCAAAATAGCAAAAGCCTCTCTCAGCCACTGGCCTGAATCTACACTGGAAGCAAC	2388
QY	1981	TTGTGCGCACCCCGCTCCCAACCTTCTTGCTGGGTAGGAGAGCTTAAAGATCACCC	2040
Db	2389	TTGTGCGCACCCCGCTCCCAACCTTCTTGCTGGGTAGGAGAGCTTAAAGATCACCC	2448
QY	2041	TAAATTTACTCATCTCTTAGTGTGCTCACAATTTGGGCTCAGCAGCTCCCGAGCACA	2100
Db	2449	TAAATTTACTCATCTCTTAGTGTGCTCACAATTTGGGCTCAGCAGCTCCCGAGCACA	2508
QY	2101	ATTTCAGGCTACCCCTCTCTTCTTGCACTGTCCCAAACTTGCTGTCTCAATTTCCGAGATC	2160
Db	2509	ATTTCAGGCTACCCCTCTCTTCTTGCACTGTCCCAAACTTGCTGTCTCAATTTCCGAGATC	2568
QY	2161	TAAATCTCCCTTACGCTCTGCGAGAAATCTTTTTCAGACCTCATAGCAACAGCCGGTTG	2220
Db	2569	TAAATCTCCCTTACGCTCTGCGAGAAATCTTTTTCAGACCTCATAGCAACAGCCGGTTG	2628
QY	2221	CTCCTTGTGAGAGAAATTTGTAGATCATTTCTCACTTCAAAATTCCTGGGCTCATCTCT	2280
Db	2629	CTCCTTGTGAGAGAAATTTGTAGATCATTTCTCACTTCAAAATTCCTGGGCTCATCTCT	2688
QY	2281	CTCATCTTGCAACCCCAACCTCTGTAAATAGATTTTACCGCATTTACGGCTGCAATCTGTAA	2340
Db	2689	CTCATCTTGCAACCCCAACCTCTGTAAATAGATTTTACCGCATTTACGGCTGCAATCTGTAA	2748
QY	2341	GTGGGATGCTCTCTTAATGAGGAGTGTTCATTTGTAATTAAGTTATTTCACTTGATAT	2400
Db	2749	GTGGGATGCTCTCTTAATGAGGAGTGTTCATTTGTAATTAAGTTATTTCACTTGATAT	2808
QY	2401	GCAATAAAGATGTGGTGGCCACTCTTTTCATGTGTGGCAGCA	2443

Db 2809 GCAATAAAGATGTGGTGCCCACTCTTTTCATGTGTGGTGCCAGCA 2851  
|||||

RESULT 12  
US-10-775-920-11  
; Sequence 11, Application US/10775920  
; Publication No. US2004017544A1  
; GENERAL INFORMATION:  
; APPLICANT: Mergen Ltd  
; TITLE OF INVENTION: POLYNUCLEOTIDE SEQUENCES AND CORRESPONDING ENCODED POLYPEPTIDES  
; TITLE OF INVENTION: OF PARTICULAR SECRETED AND MEMBRANE-BOUND PROTEINS OVEREXPRESSED  
; TITLE OF INVENTION: IN CERTAIN CANCERS  
; FILE REFERENCE: Mergen - 0010B  
; CURRENT APPLICATION NUMBER: US/10/775,920  
; CURRENT FILING DATE: 2004-02-10  
; PRIOR APPLICATION NUMBER: US 60/447,900  
; PRIOR FILING DATE: 2003-02-13  
; NUMBER OF SEQ ID NOS: 385  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 11  
; LENGTH: 2446  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-775-920-11

Query Match 99.3%; Score 2439.8; DB 9; Length 2446;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 2441; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATAACAGCATGAAGTGCCTGGAACTCGAATAGCGGTCTCTCTCCCTCGACCCCTCCCC 60  
Db 1 ATAACAGCATGAAGTGCCTGGAACTCGAATAGCGGTCTCTCTCCCTCGACCCCTCCCC 60

Qy 61 TCCTTGTCCCTCTGCTCACCCCTCGCTCGTTCCTCTCCCTCGGCGAGGCGCCTTTATA 120  
Db 61 TCCTTGTCCCTCTGCTCACCCCTCGCTCGTTCCTCTCCCTCGGCGAGGCGCCTTTATA 120

Qy 121 ACAACTCTCTCAGAGTGGCGGGGATAGTGTCTCAAGGTCTCCCCACGACTGAGGAG 180  
Db 121 ACAACTCTCTCAGAGTGGCGGGGATAGTGTCTCAAGGTCTCCCCACGACTGAGGAG 180

Qy 181 CTGCGCTGTGCCCTCTTTGCGCGGGGAAAGAGCAACAAAGTTCAAGCCCAACGCCCTTGGC 240  
Db 181 CTGCGCTGTGCCCTCTTTGCGCGGGGAAAGAGCAACAAAGTTCAAGCCCAACGCCCTTGGC 240

Qy 241 ACTAGGGTCAGATGCTACACAGTCCCTGATGGTTGCGCAATGGCCTGAATCCAA 300  
Db 241 ACTAGGGTCAGAAATGCTACACAGTCCCTGATGGTTGCGCAATGGCCTGAATCCAA 300

Qy 301 GTACTACAGACTTTGTGATAAGGTGAAGCTTGGGGCATCGTCTAGAAAACGGTGGCCAC 360  
Db 301 GTACTACAGACTTTGTGATAAGGTGAAGCTTGGGGCATCGTCTAGAAAACGGTGGCCAC 360

Qy 361 AGCGGGGTTGTGACCTCGGTGGCCTTTCAGTCTCTCCGATCTCTCCGATCTCTGCAAGGT 420  
Db 361 AGCGGGGTTGTGACCTCGGTGGCCTTTCAGTCTCTCCGATCTCTCCGATCTCTGCAAGGT 420

Qy 421 GCAGGACTCCAAAGCGGCAAAATGCTGCTACTCAGTTTCTCTTCTCTGGGTGTGT 480  
Db 421 GCAGGACTCCAAAGCGGCAAAATGCTGCTACTCAGTTTCTCTTCTCTGGGTGTGT 480

Qy 481 GGGGACTTTGGGCTCACCTTTGGCTTTTCATCATCGGACTGGAGGAGCACAGGGCCAC 540  
Db 481 GGGGACTTTGGGCTCACCTTTGGCTTTTCATCATCGGACTGGAGGAGCACAGGGCCAC 540

Qy 541 ACGCTTCTCTCTTTGGGATCTCTTTTCCATCTGCTTCTCTGCTGCTGCTCATGC 600  
Db 541 ACGCTTCTCTCTTTGGGATCTCTTTTCCATCTGCTTCTCTGCTGCTGCTCATGC 600

Qy 601 TGTGAGTCTGACCAAGCTCTGTCGGGGAGGAGAGCCCTTTTCCCTGTTGTTGATCTGGG 660  
Db 601 TGTGAGTCTGACCAAGCTCTGTCGGGGAGGAGAGCCCTTTTCCCTGTTGTTGATCTGGG 660



QY	1741	CTGGGATGACAGCGGTGAGCCACAGCTCCAGGCTAGGCCCTTAATCTTGCTGTTATTTT	1800
DB	1741		1800
QY	1801	CTGGGATGACAGCGGTGAGCCACAGCTCCAGGCTAGGCCCTTAATCTTGCTGTTATTTT	1860
DB	1801		1860
QY	1801	CCATGAGCTAAAGGCTCTGCTCATCTGAGCTCAGCTGGCTCACACAGCTCTAGGGGCTG	1860
DB	1801		1860
QY	1861	CTCTCTAACTCACAGTGGGTTTGTGAGGCTCTGTGGCCAGAGCAGACCTGCATATCT	1920
DB	1861		1920
QY	1921	GAGCAAAATAGCAAAAGCCTCTCAGCCCACTGGCTGAATCTACACTGGAAGCCAAAC	1980
DB	1921		1980
QY	1981	TTGCTGGCACCCCGCTCCCAACCTCTCTGCTGGGTAGGAGGCTAAAGATCACCC	2040
DB	1981		2040
QY	2041	TAAATTTACTCATCTCTAGTGTGCTTCAATTTGGGCTCAGCAGCTCCCCAGCACCA	2100
DB	2041		2100
QY	2101	ATTACAGTCAACCCCTCTCTTCTTGCACTGTCCCAAACTTGCTGTCATTTCCGAGATC	2160
DB	2101		2160
QY	2161	TAACTCTCCCTTACGCTCTGCCAGGAATCTTTTCAGACCTCACTAGCAAGCCGGTTG	2220
DB	2161		2220
QY	2221	CTCTTTGTGAGGAGATTTGTAGATCATTTCTCACTTTCAAATTTCTGGGCTGATCTCT	2280
DB	2221		2280
QY	2281	CTCATCTTGACCCCAACCTCTGTAAATAGATTTACCGCATTTACGGCTGCAATCTGTAA	2340
DB	2281		2340
QY	2341	GTGGGATGGTCTCCTAATGGAGGAGTGTTCATTTGATTAAGTTATTCACTCAGATAT	2400
DB	2341		2400
QY	2401	GCAATAAAGATGTGGTGGCCACTCTTTTCATGTGTGGGAGCA	2443
DB	2401		2443

RESULT 13  
US-10-224-289-3  
; Sequence 3, Application US/10224289  
; Publication No. US20030207288A1  
; GENERAL INFORMATION:  
; APPLICANT: LEWIN, DAVID A.  
; APPLICANT: STEWART, TIMOTHY A.  
; TITLE OF INVENTION: GPCR-LIKE RETINOIC ACID-INDUCED GENE 1 PROTEIN AND  
; TITLE OF INVENTION: NUCLEIC ACID  
; FILE REFERENCE: 9800081-0085  
; CURRENT APPLICATION NUMBER: US/10/224,289  
; CURRENT FILING DATE: 2002-08-20  
; PRIOR APPLICATION NUMBER: 60/313,940  
; PRIOR FILING DATE: 2001-08-20  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 2302  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-224-289-3

Query Match 93.7%; Score 2302; DB 7; Length 2302;  
Best Local Similarity 100.0%; Pred. No. 0;

		Matches 2302;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	155	CCAAGTCTCCCCAGCACTGAGGAGCTCGCTGTGCCCTCTTGCGCGGGAAGCAGC	214			
DB	1	CCAAGTCTCCCCAGCACTGAGGAGCTCGCTGTGCCCTCTTGCGCGGGAAGCAGC	60			
QY	215	ACCAAGTTCACGGCCAAAGCCCTTGGCACTAGGGTCCAGATGGCTACAACAGTCCCTGAT	274			
DB	61	ACCAAGTTCACGGCCAAAGCCCTTGGCACTAGGGTCCAGATGGCTACAACAGTCCCTGAT	120			
QY	275	GGTTCGCGCAATGGCTGAAATCCAAAGTACTACAGACTTTGTGATAAGGCTGAAGCTTGG	334			
DB	121	GGTTCGCGCAATGGCTGAAATCCAAAGTACTACAGACTTTGTGATAAGGCTGAAGCTTGG	180			
QY	335	GGCATCTGCTTAGAAACCGGTGGCCACAGCCGGGTGTGACCTCGGTGGCTTTCATGCTC	394			
DB	181	GGCATCTGCTTAGAAACCGGTGGCCACAGCCGGGTGTGACCTCGGTGGCTTTCATGCTC	240			
QY	395	ACTCTCCGATCCTCGTCTGCAAGGTGCAGGACTCAACAGCGGAAATGCTGCTACT	454			
DB	241	ACTCTCCGATCCTCGTCTGCAAGGTGCAGGACTCAACAGCGGAAATGCTGCTACT	300			
QY	455	CAGTTTCTCTTCTCCTCGGTGTGTCATCTTTGGCTCACCTTGCCCTTTCATCATC	514			
DB	301	CAGTTTCTCTTCTCCTCGGTGTGTCATCTTTGGCTCACCTTGCCCTTTCATCATC	360			
QY	515	GGACTGGAGCGGAGACAGGGCCACAGCTTCTTCTCTTTGGGATCCTCTTTTCCATC	574			
DB	361	GGACTGGAGCGGAGACAGGGCCACAGCTTCTTCTCTTTGGGATCCTCTTTTCCATC	420			
QY	575	TGCTTCTCTGCTGCTGCTCATGCTCATGCTGACCAAGCTGTCGGGGGAGGAAG	634			
DB	421	TGCTTCTCTGCTGCTGCTCATGCTCATGCTGACCAAGCTGTCGGGGGAGGAAG	480			
QY	635	CCCTTTTCTCTGCTGCTGCTCATGCTGCTGCGGCTTCCAGCTAGTCAGGATGTT	694			
DB	481	CCCTTTTCTCTGCTGCTGCTCATGCTGCTGCGGCTTCCAGCTAGTCAGGATGTT	540			
QY	695	ATCGCTATTGAATATATTGCTCTGACCATGAATAGAACCAACGTCATTTCTGAG	754			
DB	541	ATCGCTATTGAATATATTGCTCTGACCATGAATAGAACCAACGTCATTTCTGAG	600			
QY	755	CTTTCCGCTCTCTGCGCAATGAAGACTTTGCTCTGCTCACCTACGCTCTCTCTTG	814			
DB	601	CTTTCCGCTCTCTGCGCAATGAAGACTTTGCTCTGCTCACCTACGCTCTCTCTTG	660			
QY	815	ATGGCGCTGACCTTCTCATGTCCTCTTCCACTTCTGTGGTTCTTTCACGGCTGGAAG	874			
DB	661	ATGGCGCTGACCTTCTCATGTCCTCTTCCACTTCTGTGGTTCTTTCACGGCTGGAAG	720			
QY	875	AGACATGGGGCCACATCTTACCTCAGATGCTCTCTCCATTTGCCATCTGGGTGGCTGG	934			
DB	721	AGACATGGGGCCACATCTTACCTCAGATGCTCTCTCCATTTGCCATCTGGGTGGCTGG	780			
QY	935	ATCACCTCTGCTCATGCTTCTGACCTTGACCCGAGGTGGGATGACACCATCTCAGCTCC	994			
DB	781	ATCACCTCTGCTCATGCTTCTGACCTTGACCCGAGGTGGGATGACACCATCTCAGCTCC	840			
QY	995	GCCTTGGCTGCAATGGCTGGGTGTTCTGTGGCTTATGTTAGTCCCGAGTTTGGCTG	1054			
DB	841	GCCTTGGCTGCAATGGCTGGGTGTTCTGTGGCTTATGTTAGTCCCGAGTTTGGCTG	900			
QY	1055	CTCAGAAAGCAACGAAACCCCAATGATATCTGTGTAGGATGCTTTTGTAAACCTCAA	1114			
DB	901	CTCAGAAAGCAACGAAACCCCAATGATATCTGTGTAGGATGCTTTTGTAAACCTCAA	960			
QY	1115	CTCGTGAAGAGAGCTATGCTGTGGAGACAGAGCTACTCTCAAGAGAAATCACTCAA	1174			
DB	961	CTCGTGAAGAGAGCTATGCTGTGGAGAACAGAGCTACTCTCAAGAGAAATCACTCAA	1020			
QY	1175	GGTTTGAAGAGACAGGGGACACGCTCTATGCCCCCTTATTCACACATTTTTCAGCTGCAG	1234			
DB	1021	GGTTTGAAGAGACAGGGGACACGCTCTATGCCCCCTTATTCACACATTTTTCAGCTGCAG	1080			



Qy	1235	AAC	CAG	GCT	CCCC	AAAA	AGGA	AT	CT	CC	AT	CC	AC	GGG	CC	CAC	GT	TGG	CG	AG	CC	CT	TAC	1294																			
Db	1081	AAC	CAG	GCT	CCCC	AAAA	AGGA	AT	CT	CC	AT	CC	AC	GGG	CC	CAC	GT	TGG	CG	AG	CC	CT	TAC	1140																			
Qy	1295	AA	GAC	TAT	GA	GT	AA	GA	AG	GG	CAG	CT	AA	CT	CT	GT	C	T	GA	AG	T	GG	G	CA	AA	TGC	1354																
Db	1141	AA	GAC	TAT	GA	GT	AA	GA	AG	GG	CAG	CT	AA	CT	CT	GT	C	T	GA	AG	T	GG	G	CA	AA	TGC	1200																
Qy	1355	AG	CGG	CGG	CGG	CAG	AT	CT	AG	CGG	AG	CT	CA	AG	GG	AT	TG	GG	CG	GA	A	T	CT	T	GA	G	CT	CTG	1414														
Db	1201	AG	CGG	CGG	CGG	CAG	AT	CT	AG	CGG	AG	CT	CA	AG	GG	AT	TG	GG	CG	GA	A	T	CT	T	GA	G	CT	CTG	1260														
Qy	1415	AG	AAAA	CT	GT	CA	AG	CA	CT	TA	CGG	GA	CAG	TTT	GC	CT	CC	T	CC	CAG	CT	CA	AC	CA	CA	AA	TT	1474															
Db	1261	AG	AAAA	CT	GT	CA	AG	CA	CT	TA	CGG	GA	CAG	TTT	GC	CT	CC	T	CC	CAG	CT	CA	AC	CA	CA	AA	TT	1320															
Qy	1475	CT	CC	AT	GT	TGG	GG	CT	GAT	GT	GG	CT	AG	T	AG	AC	T	CC	AG	T	T	CT	T	T	AG	AG	GG	CG	CT	GT	AG	TAT	1534										
Db	1321	CT	CC	AT	GT	TGG	GG	CT	GAT	GT	GG	CT	AG	T	AG	AC	T	CC	AG	T	T	CT	T	T	AG	AG	GG	CG	CT	GT	AG	TAT	1380										
Qy	1535	TTTT	TTTT	TTTT	TG	T	CT	CAT	CC	TT	TGG	AT	CT	T	TT	TA	AG	TGG	AG	CT	C	AG	G	CA	A	CT	CA	1594															
Db	1381	TTTT	TTTT	TTTT	TG	T	CT	CAT	CC	TT	TGG	AT	CT	T	TT	TA	AG	TGG	AG	CT	C	AG	G	CA	A	CT	CA	1440															
Qy	1595	AG	TT	AG	AC	CT	TACT	CT	TTTT	TG	TT	TG	TTTT	TG	AAAA	CAG	AT	CT	TG	CT	GT	C	T	AG	C	CC	AG	1654															
Db	1441	AG	TT	AG	AC	CT	TACT	CT	TTTT	TG	TT	TG	TTTT	TG	AAAA	CAG	AT	CT	TG	CT	GT	C	T	AG	C	CC	AG	1500															
Qy	1655	G	CT	T	AG	T	G	C	A	G	T	G	C	A	G	C	C	A	G	T	G	C	A	G	C	C	A	G	1714														
Db	1501	G	CT	T	AG	T	G	C	A	G	T	G	C	A	G	C	C	A	G	T	G	C	A	G	C	C	A	G	1560														
Qy	1715	T	C	T	C	C	A	T	C	T	C	C	A	A	G	T	G	G	A	T	G	A	C	A	G	G	G	T	G	A	C	A	G	CT	C	C	AG	1774					
Db	1561	T	C	T	C	C	A	T	C	T	C	C	A	A	G	T	G	G	A	T	G	A	C	A	G	G	G	T	G	A	C	A	G	CT	C	C	AG	1620					
Qy	1775	T	A	G	C	C	T	T	A	T	C	T	T	T	T	G	T	T	T	T	T	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	1834						
Db	1621	T	A	G	C	C	T	T	A	T	C	T	T	T	T	G	T	T	T	T	T	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	1680						
Qy	1835	C	T	G	C	T	C	A	C	A	G	CT	CT	AG	G	G	C	T	GC	T	C	T	TA	AC	T	CA	C	A	G	T	G	G	G	T	T	T	G	T	AG	G	CT	CT	1894
Db	1681	C	T	G	C	T	C	A	C	A	G	CT	CT	AG	G	G	C	T	GC	T	C	T</																					

Qy	2315	ACGCATTACGGCTGCATTTCTGTAGTGGGCATGGTCTCCTTAATGAGAGTGTTCATT	2374
Db	2161	ACGCATTACGGCTGCATTTCTGTAGTGGGCATGGTCTCCTTAATGAGAGTGTTCATT	2220
Qy	2375	GTATATAAGCTTATTCACCTGAGTATGCAATAAAGATGTGGTGCCACTCTTTTCATGGTG	2434
Db	2221	GTATATAAGTATTATTCACCTGAGTATGCAATAAAGATGTGGTGCCACTCTTTTCATGGTG	2280
Qy	2435	GTGGCAGCAAAAAAAAAAAAAA 2456	
Db	2281	GTGGCAGCAAAAAAAAAAAAAA 2302	
RESULT 14			
US-10-240-425-405			
; Sequence 405, Application US/10240425			
; Publication No. US20040033502A1			
; GENERAL INFORMATION:			
; APPLICANT: Williams, Amanda			
; APPLICANT: Bolland, Joseph F.			
; APPLICANT: Lord, Reginald V.			
; APPLICANT: Alvarez, Chris			
; APPLICANT: Wetzel, Jon C.			
; APPLICANT: Scherf, Uwe			
; APPLICANT: Vockley, Joseph G.			
; TITLE OF INVENTION: Gene Expression Profiles in Esophageal Tissue			
; FILE REFERENCE: 44921-5026			
; CURRENT APPLICATION NUMBER: US/10/240,425			
; PRIOR APPLICATION NUMBER: PCT/US01/09847			
; PRIOR FILING DATE: 2002-09-30			
; PRIOR APPLICATION NUMBER: US 60/193,446			
; PRIOR FILING DATE: 2000-03-31			
; NUMBER OF SEQ ID NOS: 1588			
; SOFTWARE: PatentIn Ver. 2.1			
; SEQ ID NO 405			
; LENGTH: 2302			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
; FEATURE:			
; OTHER INFORMATION: Genbank Accession No. US20040033502A1 AF095448			
US-10-240-425-405			
Query Match 93.7%; Score 2302; DB 8; Length 2302;			
Best Local Similarity 100.0%; Pred. No. 0;			
Matches 2302; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	155	CCAAGTCTCCCCAGCACTGAGGAGTCCGCTGTCGCCCTCTTTGGCGCGGGAAGCAGC	214
Db	1	CCAAGTCTCCCCAGCACTGAGGAGTCCGCTGTCGCCCTCTTTGGCGCGGGAAGCAGC	60
Qy	215	ACCAAGTTACAGGCCAACCGCTTTGGCACTAGGGTCCAGAAATGCTACAAAGTCCCTGAT	274
Db	61	ACCAAGTTACAGGCCAACCGCTTTGGCACTAGGGTCCAGAAATGCTACAAAGTCCCTGAT	120
Qy	275	GGTTGCGCAATGGCCTGAAATCCAAGTACTACAGACTTTTGTGATAGGCTGAAGCTTGG	334
Db	121	GGTTGCGCAATGGCCTGAAATCCAAGTACTACAGACTTTTGTGATAGGCTGAAGCTTGG	180
Qy	335	GGCATCGTCTAGAAACGGTGGCCACAGCGGGTTGTGACCTCGGTGGCTTCATGCTC	394
Db	181	GGCATCGTCTAGAAACGGTGGCCACAGCGGGTTGTGACCTCGGTGGCTTCATGCTC	240
Qy	395	ACTCTCCGATCTCGTCTCAGAGTGCAGGACTCCAAACAGCGCAAAAATGCTGCCTACT	454
Db	241	ACTCTCCGATCTCGTCTCAGAGTGCAGGACTCCAAACAGCGCAAAAATGCTGCCTACT	300
Qy	455	CAGTTTCTCTCTCTCGTGGTGTGGGCATCTTTGGCCTCACCTTCGCTTCATCATC	514
Db	301	CAGTTTCTCTCTCTCGTGGTGTGGGCATCTTTGGCCTCACCTTCGCTTCATCATC	360
Qy	515	GGACTGGAAGGAGCAAGGGCCCAACACGCTTCTTCTTTGGGATTCCTTTTTCATC	574

Db 361 GGACTGGAGGAGCACAGGGCCCCACACGCTTCTCTCTTTGGGANTCTCTTTTCCATC 420  
Qy 575 TGCTTTCTCCTGCTGCTGCTCATGTGTCAGTCTGACCAAGCTCGTCCGGGGGAGGAG 634  
Db 421 TGCTTCTCCTGCTGCTGCTCATGTGTCAGTCTGACCAAGCTCGTCCGGGGGAGGAG 480  
Qy 635 CCCCCTTCCCTTTGGTGATTCCTGGGTCGCGGCTGAGGCTTCAAGCTAGTCAGAGATGT 694  
Db 481 CCCCCTTCCCTTTGGTGATTCCTGGGTCGCGGCTGAGGCTTCAAGCTAGTCAGAGATGT 540  
Qy 695 ATCGCTATGATATATGTCCTGACCATGATAGGACCAAGCTCAATGCTCTTTTCTGAG 754  
Db 541 ATCGCTATGATATATGTCCTGACCATGATAGGACCAAGCTCAATGCTCTTTTCTGAG 600  
Qy 755 CTTTCCGCTCCTCGTCCGAATGAAGACTTTGTCCTGCTCAGCTCAGCTCCTCTTTCTG 814  
Db 601 CTTTCCGCTCCTCGTCCGAATGAAGACTTTGTCCTGCTCAGCTCAGCTCCTCTTTCTG 660  
Qy 815 ATGGCGCTGACCTTCTCATGTCCTCTTCACTTCTGCTGCTTCTTCAAGGGCTGGAAG 874  
Db 661 ATGGCGCTGACCTTCTCATGTCCTCTTCACTTCTGCTGCTTCTTCAAGGGCTGGAAG 720  
Qy 875 AGACATGGGGCCCCACATCTACCTCAGCATGCTCTCTCCATTCGCTCTGGGTCGCTGG 934  
Db 721 AGACATGGGGCCCCACATCTACCTCAGCATGCTCTCTCCATTCGCTCTGGGTCGCTGG 780  
Qy 935 ATCACCTGCTCATGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 994  
Db 781 ATCACCTGCTCATGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840  
Qy 995 GCGTTGGCTGCGCAATGGCTGGGTGCTTCTGTTGGGCTTATGTTAGTCCGAGTTTGGGTG 1054  
Db 841 GCGTTGGCTGCGCAATGGCTGGGTGCTTCTGTTGGGCTTATGTTAGTCCGAGTTTGGGTG 900  
Qy 1055 CTCACAAAGCAAGAAACCCATGATATCTGTTGAGGATGCTTCTGTAACCTCAA 1114  
Db 901 CTCACAAAGCAAGAAACCCATGATATCTGTTGAGGATGCTTCTGTAACCTCAA 960  
Qy 1115 CTCGTGAAGAGAGCTATGTTGGAGAACAGAGCTTCTCTCAAGAGAAATCACTCAA 1174  
Db 961 CTCGTGAAGAGAGCTATGTTGGAGAACAGAGCTTCTCTCAAGAGAAATCACTCAA 1020  
Qy 1175 GGTTTGAAGAGACAGGGGACAGCTCTATGCCCCCTATTCACACATTTTCAGTGCAG 1234  
Db 1021 GGTTTGAAGAGACAGGGGACAGCTCTATGCCCCCTATTCACACATTTTCAGTGCAG 1080  
Qy 1235 AACCAAGCTCCCAAAAGGAATTCCTCAACAGGGCCAGCTTGGCCGAGCCCTTAC 1294  
Db 1081 AACCAAGCTCCCAAAAGGAATTCCTCAACAGGGCCAGCTTGGCCGAGCCCTTAC 1140  
Qy 1295 AAAGACTATGAAGTAAAGAAAGAGGACAGCTAACTCTGTCTGAAGAGTGGGACAAATGC 1354  
Db 1141 AAAGACTATGAAGTAAAGAAAGAGGACAGCTAACTCTGTCTGAAGAGTGGGACAAATGC 1200  
Qy 1355 AGCCGGGGGACAGATCTAGCGGAGCTCAAGGGATGCGGGAATCTTGAGTCTTCTG 1414  
Db 1201 AGCCGGGGGACAGATCTAGCGGAGCTCAAGGGATGCGGGAATCTTGAGTCTTCTG 1260  
Qy 1415 AGAAACTGTACAAGACACTAGCGGAACAGTTTGGCTCCCTCCAGCTCAACCAAAAT 1474  
Db 1261 AGAAACTGTACAAGACACTAGCGGAACAGTTTGGCTCCCTCCAGCTCAACCAAAAT 1320  
Qy 1475 CTTCCATGCTGGGCTGATGCGGCTAGTAAGACTCCAGTCTTGAAGCGCTGTAGTAT 1534  
Db 1321 CTTCCATGCTGGGCTGATGCGGCTAGTAAGACTCCAGTCTTGAAGCGCTGTAGTAT 1380  
Qy 1535 TTTTTTTTTTGTCTCATCTTGGATCTTTTAAAGTGGAGTCTCAGGCACTCA 1594  
Db 1381 TTTTTTTTTTGTCTCATCTTGGATCTTTTAAAGTGGAGTCTCAGGCACTCA 1440  
Qy 1595 AGTTTAGACCTTACTCTTTTGTGTTTGTGTTTGAACAGGATCTTGCTCTGCAACCCAG 1654  
Db 1441 AGTTTAGACCTTACTCTTTTGTGTTTGTGTTTGAACAGGATCTTGCTCTGCAACCCAG 1500

RESULT 15

US-10-775-920-9

; Sequence 9, Application US/10775920

; Publication No. US2004017544A1

; GENERAL INFORMATION:

; APPLICANT: Mergen Ltd

; TITLE OF INVENTION: POLYNUCLEOTIDE SEQUENCES AND CORRESPONDING ENCODED POLYPEPTIDES

; TITLE OF INVENTION: OF PARTICULAR SECRETED AND MEMBRANE-BOUND PROTEINS OVEREXPRESSED

; TITLE OF INVENTION: IN CERTAIN CANCERS

; FILE REFERENCE: Mergen - 0010B

; CURRENT APPLICATION NUMBER: US/10/775,920

; CURRENT FILING DATE: 2004-02-10

; PRIOR APPLICATION NUMBER: US 60/447,900

; PRIOR FILING DATE: 2003-02-13

; NUMBER OF SEQ ID NOS: 385



Qy	2195	AGACCTCTACTAGCACAAAGCCCGGTGGCTCCTTGTGTCAGGAGAAATTTGTAGATCAATTCAC	2254
Db	2041	AGACCTCTACTAGCACAAAGCCCGGTGGCTCCTTGTGTCAGGAGAAATTTGTAGATCAATTCAC	2100
Qy	2255	TTCAAATTCCTGGGCTGATACCTCTCTCATCTTGCACCCCAACCTCTGTAAATAGATT	2314
Db	2101	TTCAAATTCCTGGGCTGATACCTCTCTCATCTTGCACCCCAACCTCTGTAAATAGATT	2160
Qy	2315	ACCGCATTTACGGCTGCAATTCCTGAAGTGGGCATGGTCTCTTAATGGAGAGTGTTCATT	2374
Db	2161	ACCGCATTTACGGCTGCAATTCCTGAAGTGGGCATGGTCTCTTAATGGAGAGTGTTCATT	2220
Qy	2375	GTATAATAAGTTATTACCTGAGTATGCAATAAAGATGTGGCCACTCTTTTCATGGTG	2434
Db	2221	GTATAATAAGTTATTACCTGAGTATGCAATAAAGATGTGGCCACTCTTTTCATGGTG	2280
Qy	2435	GTGGCAGCAAAAAAAAAAAAAA	2456
Db	2281	GTGGCAGCAAAAAAAAAAAAAA	2302

Search completed: June 10, 2006, 15:59:14  
Job time : 3009 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2006, 14:55:02 ; Search time 1403 Seconds

(without alignments)  
12205.162 Million cell updates/sec

Title: US-10-600-816-2

Perfect score: 2456

Sequence: 1 ataacagcatgaagtgcggt.....ggcagcaaaaaaaaaaaaaa 2456

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 5244920 seqs, 3486124231 residues

Total number of hits satisfying chosen parameters: 10489840

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_8:\*

1: Geneseqn1980s:\*

2: Geneseqn1990s:\*

3: Geneseqn2000s:\*

4: Geneseqn2001as:\*

5: Geneseqn2001bs:\*

6: Geneseqn2002as:\*

7: Geneseqn2002bs:\*

8: Geneseqn2003as:\*

9: Geneseqn2003bs:\*

10: Geneseqn2003cs:\*

11: Geneseqn2003ds:\*

12: Geneseqn2004as:\*

13: Geneseqn2004bs:\*

14: Geneseqn2005as:\*

15: Geneseqn2005bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2456	100.0	2456	8	ABZ42832 Human G p
2	2456	100.0	2456	11	ADN39301 Cancer/an
3	2456	100.0	2456	12	ADI28459 Human GPC
4	2456	100.0	2456	13	ADR48221 Human ret
5	2456	100.0	2456	13	ACN39388 Tumour-as
6	2456	100.0	2456	13	ADU06126 Novel bro
7	2456	100.0	2456	14	AEA00088 Human TAT
8	2456	100.0	2456	14	AEA00508 Human TAT
9	2456	100.0	2456	14	AED47480 Retinoic
10	2448	99.7	2456	12	ADI28525 Human GPC
11	2443	99.5	4239	11	ACN89274 Breast ca
12	2443	99.5	6730	13	ADK97494 Pancreati
13	2441.6	99.4	2456	10	ACC72695 Human can
14	2439.8	99.3	2446	4	AHL14688 Human cdn
15	2437	99.2	2456	13	ADR43992 Human bre
16	2302	93.7	2302	6	ABT10173 Human bre
17	2302	93.7	2302	8	ACC58386 Human GPC
17	2302	93.7	2302	10	ADD93240 RAIGI cod

19	2302	93.7	2302	10	ADL24773
20	2288.8	93.2	3371	4	AAH72766
21	2286	93.1	2316	10	ABT31923
22	2260.4	92.0	2593	6	ABQ54954
23	1701.2	69.3	1718	3	AAZ90056
24	1601.8	65.2	1619	4	AAF58615
25	1441.4	58.7	1460	14	AEA00112
26	1441.4	58.7	1460	14	AEA00632
27	1177.6	47.9	2347	14	ADZ88660
28	1126	45.8	1212	3	AAZ61776
29	1126	45.8	1212	4	AAC99709
30	1126	45.8	1212	6	ABL34861
31	1082.6	44.1	1114	2	AAZ97687
32	1082.6	44.1	1114	12	ADP18954
33	1074	43.7	1074	12	ADO30035
34	1067.8	43.5	1788	10	ADF70573
35	1031.6	42.0	1228	8	ACA04775
36	932.6	38.0	1000	4	AAI21778
37	932.6	38.0	1000	4	ABA66850
38	932.6	38.0	1000	4	AAI47061
39	932.6	38.0	1000	4	ABA48935
40	932.6	38.0	1000	4	ABA33918
41	932.6	38.0	1000	4	AAK41008
42	932.6	38.0	1000	4	AAK15285
43	932.6	38.0	1000	4	ABS40600
44	932.6	38.0	1000	5	AAI07463
45	932.6	38.0	1000	6	ABS14980

## ALIGNMENTS

### RESULT 1

ABZ42832

ID ABZ42832 standard; DNA; 2456 BP.

XX

AC ABZ42832;

XX

DT 04-MAR-2003 (first entry)

XX

DE Human G protein-coupled receptor RAIGI nucleotide SEQ ID NO:453.

XX

KW G protein-coupled receptor; GPCR; antigenic peptide; gene therapy;

KW G protein-coupled receptor modulator; antibody; immune-related disease;

KW growth-related disease; cell regeneration-related disease; AIDS; cancer;

KW immunological-related cell proliferative disease; autoimmune disease;

KW Alzheimer's disease; atherosclerosis; infection; osteoarthritis; allergy;

KW osteoporosis; cardiomyopathy; inflammation; Crohn's disease; diabetes;

KW graft versus host disease; Parkinson's disease; multiple sclerosis; pain;

KW psoriasis; anxiety; depression; schizophrenia; dementia; memory loss;

KW mental retardation; epilepsy; asthma; tuberculosis; obesity; trauma;

KW hypertension; hypotension; renal disorder; rheumatoid arthritis; nausea;

KW ulcer; gene; ds.

XX

XX Homo sapiens.

XX

PN WO200261087-A2.

XX

PD 08-AUG-2002.

XX

PF 19-DEC-2001; 2001WO-US050107.

XX

PR 19-DEC-2000; 2000US-0257144P.

XX

PA (LIFE-) LIFESPAN BIOSCIENCES INC.

XX

PI Burner GC, Roush CL, Brown JF;

XX

DR WPI; 2003-046718/04.

DR

DR P-PSDB; ABP81984.

XX

PT New isolated antigenic peptides e.g., for G protein-coupled receptors

PT (GPCR), useful for diagnosing and designing drugs for treating conditions

PT in which GPCRs are involved, e.g. AIDS, Alzheimer's disease, cancer or  
PT autoimmune diseases.

PS Disclosure; Fig 1; 523pp; English.

XX  
XX  
CC The present invention describes antigenic peptides (I) comprising: (a)  
CC any one of 1601 sequences (see ABP82019 to ABP83619) of 12-24 amino  
CC acids. Also described: (1) an assay for the detection of a particular G  
CC protein-coupled receptor (GPCR) or a candidate polypeptide in a sample;  
CC and (2) an isolated antibody having high specificity and high affinity or  
CC avidity for a particular GPCR. (I) can be used as GPCR modulators and in  
CC gene therapy. The antigenic peptides for GPCRs are useful in detecting an  
CC antibody against a particular GPCR, and in the production of specific  
CC antibodies. The peptides and antibodies are also useful for detecting the  
CC presence or absence of corresponding GPCRs. The antigenic peptides for  
CC GPCRs and antibodies are useful for diagnosing and designing drugs for  
CC treating immune-related diseases, growth-related diseases, cell  
CC regeneration-related diseases, immunological-related cell proliferative  
CC diseases, or autoimmune diseases, e.g. AIDS, Alzheimer's disease,  
CC atherosclerosis, bacterial, fungal, protozoan or viral infections,  
CC osteoarthritis, osteoporosis, cancer, cardiomyopathy, chronic and acute  
CC inflammation, allergies, Crohn's disease, diabetes, graft versus host  
CC disease, Parkinson's disease, multiple sclerosis, pain, psoriasis,  
CC anxiety, depression, schizophrenia, dementia, mental retardation, memory  
CC loss, epilepsy, asthma, tuberculosis, obesity, nausea, hypertension,  
CC hypotension, renal disorders, rheumatoid arthritis, trauma, ulcers, or  
CC any other disorder in which GPCRs are involved. The antibodies may be  
CC used in immunoassays and immunodiagnosis. AB242523 to AB242869 encode  
CC GPCR proteins given in ABP81675 to ABP82018, which are used in the  
CC exemplification of the present invention

XX  
SQ Sequence 2456 BP; 519 A; 722 C; 571 G; 644 T; 0 U; 0 Other;

Query Match 100.0%; Score 2456; DB 8; Length 2456;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATAAAGCATGAAGTGGCGTGAATGGAATAGGCGTGTCTCTCCCTCGACCCCTCCCCC 60  
DB 1 ATAAAGCATGAAGTGGCGTGAATGGAATAGGCGTGTCTCTCCCTCGACCCCTCCCCC 60  
QY 61 TCCTTGTCCCTGTCTACCCCTCGCTGTTCCCTCCCTCGCGAGGCGCCCTTTATA 120  
DB 61 TCCTTGTCCCTGTCTACCCCTCGCTGTTCCCTCCCTCGCGAGGCGCCCTTTATA 120  
QY 121 ACAACTGTCTCAGAGTGCAGGCGGGATAGCTGTCTCAAGGTCTCCCTCCAGCACTGAGGAG 180  
DB 121 ACAACTGTCTCAGAGTGCAGGCGGGATAGCTGTCTCAAGGTCTCCCTCCAGCACTGAGGAG 180  
QY 181 CTGCGCTGTCGCTCTTTGCGCGGGAGAGCAAGTTCAGGCTTCAAGTTCAGCGCCAGCTTGGC 240  
DB 181 CTGCGCTGTCGCTCTTTGCGCGGGAGAGCAAGTTCAGGCTTCAAGTTCAGCGCCAGCTTGGC 240  
QY 241 ACTAGGGTCCAGAAAGGCTTACAAAGTCCCTGATGTTGCGCAATGGCTGCAATCCAA 300  
DB 241 ACTAGGGTCCAGAAAGGCTTACAAAGTCCCTGATGTTGCGCAATGGCTGCAATCCAA 300  
QY 301 GTACTACAGACTTGTGTGAAGGCTTGAAGCTTGGGGCATCGTCTAGAAACGGTGGCCAC 360  
DB 301 GTACTACAGACTTGTGTGAAGGCTTGAAGCTTGGGGCATCGTCTAGAAACGGTGGCCAC 360  
QY 361 AGCCGGGTGTGACCTCGGTGGCCCTTATGCTCACTCTCCGATCTCTCGTCTGCAAGGT 420  
DB 361 AGCCGGGTGTGACCTCGGTGGCCCTTATGCTCACTCTCCGATCTCTCGTCTGCAAGGT 420  
QY 421 GCAGACTCCAAAGCGGAAATGCTGCTACTCAGTTCTCTCTCCCTGGGTGTGT 480  
DB 421 GCAGACTCCAAAGCGGAAATGCTGCTACTCAGTTCTCTCTCCCTGGGTGTGT 480  
QY 481 GGGCATCTTTGGCCCTCACCTTCGCTTTCATCATCGGACTGGAGCGGAGCACAGGCCCCAC 540  
DB 481 GGGCATCTTTGGCCCTCACCTTCGCTTTCATCATCGGACTGGAGCGGAGCACAGGCCCCAC 540

QY 541 AGCTTCTTCTCTTTGGGATCCTCTTTTCCATCTGCTTCTCTGCTGCTGCTGCTCATGC 600  
DB 541 AGCTTCTTCTCTTTGGGATCCTCTTTTCCATCTGCTTCTCTGCTGCTGCTGCTCATGC 600  
QY 601 TGTGAGTCTGACCAAGCTCGTCCGGGGAGGAGCCCTTTTCCCTGTTGGTGAATCTCGGG 660  
DB 601 TGTGAGTCTGACCAAGCTCGTCCGGGGAGGAGCCCTTTTCCCTGTTGGTGAATCTCGGG 660  
QY 661 TGTGAGTCTGAGCTTTCAGCTAGTCTCAGGATGTTATCGCTATTGAATATTGTTCTGAC 720  
DB 661 TGTGAGTCTGAGCTTTCAGCTAGTCTCAGGATGTTATCGCTATTGAATATTGTTCTGAC 720  
QY 721 CATGAATAGGACCAAGCTCAATGCTTTTCTGAGCTTTCGCTCTCTCTGCGCAATGAAGA 780  
DB 721 CATGAATAGGACCAAGCTCAATGCTTTTCTGAGCTTTCGCTCTCTCTGCGCAATGAAGA 780  
QY 781 CTCTGCTCTCTGCTCACTAGCTCTCTCTTGTGATGGCTGACCTTCTCTCATGTCCTC 840  
DB 781 CTCTGCTCTCTGCTCACTAGCTCTCTCTTGTGATGGCTGACCTTCTCTCATGTCCTC 840  
QY 841 CTCTACCTTCTGTGTTCTCTTCAACGGGCTGGAAGACATGGGGCCCAATCTACTCTAC 900  
DB 841 CTCTACCTTCTGTGTTCTCTTCAACGGGCTGGAAGACATGGGGCCCAATCTACTCTAC 900  
QY 901 GATGCTCTCTCATTTGCACTCTGGTGGCTTGGATCACCCTGCTCATGCTTCTCTGACTT 960  
DB 901 GATGCTCTCTCATTTGCACTCTGGTGGCTTGGATCACCCTGCTCATGCTTCTCTGACTT 960  
QY 961 TGACCGCAGGTGGGATGACACCATCTCAGCTCTCCCTTGGCTGCAATGGCTGGGTGT 1020  
DB 961 TGACCGCAGGTGGGATGACACCATCTCAGCTCTCCCTTGGCTGCAATGGCTGGGTGT 1020  
QY 1021 CTGTTGGCTTATGTTAGTCCGAGTTTGGCTGTCTCAAAAGCAACGAAACCCCATGA 1080  
DB 1021 CTGTTGGCTTATGTTAGTCCGAGTTTGGCTGTCTCAAAAGCAACGAAACCCCATGA 1080  
QY 1081 TTATCTGTTGAGGATGCTTCTGTAAACCTCACTCGTGAAGAGACTATGTTGTGA 1140  
DB 1081 TTATCTGTTGAGGATGCTTCTGTAAACCTCACTCGTGAAGAGACTATGTTGTGA 1140  
QY 1141 GAAACAGAGCTACTCTCAAGAGGAAATCACTCAAGTTTGAAGAGACAGGGGACACGCT 1200  
DB 1141 GAAACAGAGCTACTCTCAAGAGGAAATCACTCAAGTTTGAAGAGACAGGGGACACGCT 1200  
QY 1201 CTATGCCCCCTTATTCACACATTTTTCAGTGTGAGAACAGGCTTCCCAAAAGAAATTTCT 1260  
DB 1201 CTATGCCCCCTTATTCACACATTTTTCAGTGTGAGAACAGGCTTCCCAAAAGAAATTTCT 1260  
QY 1261 CATCCACGGGCCCCGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAGAAAGAGGG 1320  
DB 1261 CATCCACGGGCCCCGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAGAAAGAGGG 1320  
QY 1321 CAGCTAACTCTGCTCTGAAGAGTGGGACAAATGACGCGGGCGGAGACTTAGCGGGAGC 1380  
DB 1321 CAGCTAACTCTGCTCTGAAGAGTGGGACAAATGACGCGGGCGGAGACTTAGCGGGAGC 1380  
QY 1381 TCAAGAGGATGCGGCGAAATCTTGAGTCTTCTGAGAAAACTGTACAGACACTACAGGGA 1440  
DB 1381 TCAAGAGGATGCGGCGAAATCTTGAGTCTTCTGAGAAAACTGTACAGACACTACAGGGA 1440  
QY 1441 ACAGTTTCCCTCCCTCCAGGCTCAACCAAACTTCTCCATGCTGGGCTGATGTGGCT 1500  
DB 1441 ACAGTTTCCCTCCCTCCAGGCTCAACCAAACTTCTCCATGCTGGGCTGATGTGGCT 1500  
QY 1501 AGTAAGACTCCAGTTCTTAGAGGGCTGTAGTATTTTTTTTTTTTGTCTCATCTTTGG 1560  
DB 1501 AGTAAGACTCCAGTTCTTAGAGGGCTGTAGTATTTTTTTTTTTTGTCTCATCTTTGG 1560  
QY 1561 ATACTCTTTTAAAGTGGAGTCTAGGCAACTCAAGTTTAGACCTTACTCTTTTGTGTT 1620  
DB 1561 ATACTCTTTTAAAGTGGAGTCTCAGGCAACTCAAGTTTAGACCTTACTCTTTTGTGTT 1620  
QY 1621 GTTTTTTGAACAGGATCTTGTCTGTCTCAACCCAGGCTTTGAGTGCAGTGTGCGATCACAG 1680



QY	121	ACAACTGCTCAGAGTCGAGGCGGGATAGCTGTCCAAAGTCTCCCCAGCACTGAGGAG	180
Db	121	ACAACTGCTCAGAGTCGAGGCGGGATAGCTGTCCAAAGTCTCCCCAGCACTGAGGAG	180
QY	181	CTCGCCTGCTGCCCTCTTCGCGCGGGGAAGCAGCACCAGATTACGGCCAAAGCCTTGGC	240
Db	181	CTCGCCTGCTGCCCTCTTCGCGCGGGGAAGCAGCACCAGATTACGGCCAAAGCCTTGGC	240
QY	241	ACTAGGGTCCAGAAATGGCTTACAACAGTCCCTGATGGTTCCCGCAATGGCCTGAAATCCAA	300
Db	241	ACTAGGGTCCAGAAATGGCTTACAACAGTCCCTGATGGTTCCCGCAATGGCCTGAAATCCAA	300
QY	301	GTACTACAGACTTTGTGATPAAGGCTGAAGCTTTGGGCACTCGTCTTAGAAAAAGGTGGCCAC	360
Db	301	GTACTACAGACTTTGTGATPAAGGCTGAAGCTTTGGGCACTCGTCTTAGAAAAAGGTGGCCAC	360
QY	361	AGCCGGGTGTGACCTCGGTGGCCTTCATGCTCACTCTCCGATCCTCGTCTGCAAGT	420
Db	361	AGCCGGGTGTGACCTCGGTGGCCTTCATGCTCACTCTCCGATCCTCGTCTGCAAGT	420
QY	421	GCAGGACTCCAAACAGCGCAAAAATGCTGCTACTCAGTTTCTCTTCTCCTCGGTGTGT	480
Db	421	GCAGGACTCCAAACAGCGCAAAAATGCTGCTACTCAGTTTCTCTTCTCCTCGGTGTGT	480
QY	481	GGGCATCTTTGGCCTCACTTCCGCTTCATATCGGACTGGAAGGAGCAGAGGCCAC	540
Db	481	GGGCATCTTTGGCCTCACTTCCGCTTCATATCGGACTGGAAGGAGCAGAGGCCAC	540
QY	541	AGCCTTCTCTCTTTGGGATCCTCTTTTCCATCTGCTTCTCTCGCTGCTGCTCATGC	600
Db	541	AGCCTTCTCTCTTTGGGATCCTCTTTTCCATCTGCTTCTCTCGCTGCTGCTCATGC	600
QY	601	TGTGAGTCTGACCAAGCTCGTCGGGGAGGAAGCCCTTTCCCTGTTGGTGAATCTGGG	660
Db	601	TGTGAGTCTGACCAAGCTCGTCGGGGAGGAAGCCCTTTCCCTGTTGGTGAATCTGGG	660
QY	661	CTGCGCGTGGGCTTCAAGCTAGTCAGGATGTATCGCTATGAAATATATTGCTCTGAC	720
Db	661	CTGCGCGTGGGCTTCAAGCTAGTCAGGATGTATCGCTATGAAATATATTGCTCTGAC	720
QY	721	CATGAATAGGACCAAGTCAATGTCTTTTCTGAGCTTCCGCTCCCTCGCGCAATGAAGA	780
Db	721	CATGAATAGGACCAAGTCAATGTCTTTTCTGAGCTTCCGCTCCCTCGCGCAATGAAGA	780
QY	781	CTTTGTCTCTCTGCTCACCCTAGCTCTCTTTCTTGATGGCGCTGACCTTCTCATGTCTC	840
Db	781	CTTTGTCTCTCTGCTCACCCTAGCTCTCTTTCTTGATGGCGCTGACCTTCTCATGTCTC	840
QY	841	CTTCACTTCTGTGGTTCCTTCAAGGCTGGAAGAGACATGGGGCCCAATCTACCTCAC	900
Db	841	CTTCACTTCTGTGGTTCCTTCAAGGCTGGAAGAGACATGGGGCCCAATCTACCTCAC	900
QY	901	GATGCTCTCTCCATGCCATCTGGGTGGCTGGATCACCCTGCTCATGCTTCTGACTT	960
Db	901	GATGCTCTCTCCATGCCATCTGGGTGGCTGGATCACCCTGCTCATGCTTCTGACTT	960
QY	961	TGACCGCAGGTGGGATGACACCATCTCAGCTCCGCTTGGCTGCGCAATGGCTGGGTGT	1020
Db	961	TGACCGCAGGTGGGATGACACCATCTCAGCTCCGCTTGGCTGCGCAATGGCTGGGTGT	1020
QY	1021	CCTGTGGCTTATGTTAGTCCGAGTTTGGCTGTCTCAAAAGCAACGAAACCCCATGGA	1080
Db	1021	CCTGTGGCTTATGTTAGTCCGAGTTTGGCTGTCTCAAAAGCAACGAAACCCCATGGA	1080
QY	1081	TTATCTGTGTGAGGATGCTTTCTGTAACCTCAACTCGTGAAGAGAGCTATGGTGTGA	1140
Db	1081	TTATCTGTGTGAGGATGCTTTCTGTAACCTCAACTCGTGAAGAGAGCTATGGTGTGA	1140
QY	1141	GAAACAGAGCTTCTCTCAAGAGGAAATCACTCAAGGTTTTGAAGACAGAGGGACACGCT	1200
Db	1141	GAAACAGAGCTTCTCTCAAGAGGAAATCACTCAAGGTTTTGAAGACAGAGGGACACGCT	1200

QY	1201	CTATGCCCTTATTTCCACACATTTTTCAGCTGCAGAACCAAGCTCTCCCAAAAAGAAATCTTC	1260
Db	1201	CTATGCCCTTATTTCCACACATTTTTCAGCTGCAGAACCAAGCTCTCCCAAAAAGAAATCTTC	1260
QY	1261	CATCCACAGGGCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAGAAAGAGGG	1320
Db	1261	CATCCACAGGGCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAGAAAGAGGG	1320
QY	1321	CAGCTAACTCTGCTCTGAAGAGTGGGACAAATGACGCGGGCGGAGATCTTAGCGGAGC	1380
Db	1321	CAGCTAACTCTGCTCTGAAGAGTGGGACAAATGACGCGGGCGGAGATCTTAGCGGAGC	1380
QY	1381	TCAAAAGGATGTGGCGGAAATCTTGAGTCTTCTGAGAAAACCTGTACAAGACACTACGGGA	1440
Db	1381	TCAAAAGGATGTGGCGGAAATCTTGAGTCTTCTGAGAAAACCTGTACAAGACACTACGGGA	1440
QY	1441	ACAGTTTGCCTCCCTCCAGCCTCAACCAAACTTCTTCCATCTGGGGCTGATGGGGCT	1500
Db	1441	ACAGTTTGCCTCCCTCCAGCCTCAACCAAACTTCTTCCATCTGGGGCTGATGGGGCT	1500
QY	1501	AGTAAGACTCCAGTTCTTAGAGGCGCTGTAGTAATTTTTTTTTTGTCTCATCTCTTTGG	1560
Db	1501	AGTAAGACTCCAGTTCTTAGAGGCGCTGTAGTAATTTTTTTTTTGTCTCATCTCTTTGG	1560
QY	1561	ATACTTCTTTAAGTGGGAGTCTCAGGCAACTCAAGTTTAGACCCCTTACTCTTTTGT	1620
Db	1561	ATACTTCTTTAAGTGGGAGTCTCAGGCAACTCAAGTTTAGACCCCTTACTCTTTTGT	1620
QY	1621	GTTTTTTGAACAGGATCTTGTCTGTCAACCAGGCTTGAGTGCAGTGGCGATCACAG	1680
Db	1621	GTTTTTTGAACAGGATCTTGTCTGTCAACCAGGCTTGAGTGCAGTGGCGATCACAG	1680
QY	1681	CCAGTGCAGCCTCGACCACTGTGCTCAAGCAATCTCCCATCTCCATCTCCCAAAGTG	1740
Db	1681	CCAGTGCAGCCTCGACCACTGTGCTCAAGCAATCTCCCATCTCCATCTCCCAAAGTG	1740
QY	1741	CTGGGATGACAGGCGTGAGCCACAGCTCCAGCCTTAGGCCCTTAATCTTGCTGTATTTT	1800
Db	1741	CTGGGATGACAGGCGTGAGCCACAGCTCCAGCCTTAGGCCCTTAATCTTGCTGTATTTT	1800
QY	1801	CCATGGACTAAAGTCTGGTCTATCTGAGCTCAGCTGGCTCACACAGCTCTTAGGGCCCTG	1860
Db	1801	CCATGGACTAAAGTCTGGTCTATCTGAGCTCAGCTGGCTCACACAGCTCTTAGGGCCCTG	1860
QY	1861	CTCCTCTAACTCACAGTGGGTTTTGTGAGGCTCTGTGGCCAGAGCAGACCTGCATATCT	1920
Db	1861	CTCCTCTAACTCACAGTGGGTTTTGTGAGGCTCTGTGGCCAGAGCAGACCTGCATATCT	1920
QY	1921	GAGCAAAATAGCAAAAGCCTCTCTCAGCCACTGGCCTGAATCTACACTGGAAGCCAAAC	1980
Db	1921	GAGCAAAATAGCAAAAGCCTCTCTCAGCCACTGGCCTGAATCTACACTGGAAGCCAAAC	1980
QY	1981	TTGCTGGCACCCCGCTCCCCAAACCTTCTTCCCTGGGTAGGAGAGGCTTAAAGATCACCC	2040
Db	1981	TTGCTGGCACCCCGCTCCCCAAACCTTCTTCCCTGGGTAGGAGAGGCTTAAAGATCACCC	2040
QY	2041	TAAATTTACTCATCTCTTAGTGTGCCTCAATTTGGGCGCTCAGCAGCTCCCGAGCACCA	2100
Db	2041	TAAATTTACTCATCTCTTAGTGTGCCTCAATTTGGGCGCTCAGCAGCTCCCGAGCACCA	2100
QY	2101	ATTTCAGGTCACCCCTCTCTTTTGGCATGTGCCCAACTTGTCTCAATTTCCGAGATC	2160
Db	2101	ATTTCAGGTCACCCCTCTCTTTTGGCATGTGCCCAACTTGTCTCAATTTCCGAGATC	2160
QY	2161	TAAATCTCCCTTACGCTCTGCGAGGAAATTTCTTTAGACCTCTACTAGCAAGCCCGTTG	2220
Db	2161	TAAATCTCCCTTACGCTCTGCGAGGAAATTTCTTTAGACCTCTACTAGCAAGCCCGTTG	2220
QY	2221	CTCCTTGTGAGAGAAATTTGTAGATCAATCTCACTTCAAAATTTCTGGGGCTGATCTCT	2280
Db	2221	CTCCTTGTGAGAGAAATTTGTAGATCAATCTCACTTCAAAATTTCTGGGGCTGATCTCT	2280
QY	2281	CTCATCTTGCACCCCAACCTCTGTAAATAGATTTTACCGCATTTTACGGCTGCTTCTGTAA	2340



```

Db      2281  CTATCTTGCACCCCAACCTCTGTAATAGATTACCGCATTTACGGCTGCTCTGTAA 2340
QY      2341  GTGGGATGCTCTCTTAATGGAGAGTGTTTCATGTATATATATAGTTATTCACCTGAGTAT 2400
Db      2341  GTGGGATGCTCTCTTAATGGAGAGTGTTTCATGTATATATATAGTTATTCACCTGAGTAT 2400
QY      2401  GCAATAAAGATGTTGGTGGCCACTCTTTTCATGTGTTGGTGGCAGCAAAAAA 2456
Db      2401  GCAATAAAGATGTTGGTGGCCACTCTTTTCATGTGTTGGTGGCAGCAAAAAA 2456

RESULT 3
ADI28459
ID      ADI28459 standard; cDNA; 2456 BP.
XX
AC      ADI28459;
XX
XX      22-APR-2004 (first entry)
XX
DE      Human GPCR retinoic acid induced 3 (RAI-3) cDNA.
XX
KW      Retinoic acid induced 3; RAI-3; human; G-protein coupled receptor; GPCR;
KW      antiinflammatory; immunosuppressive; cytostatic; cardiant; antiallergic;
KW      broncholytic; gene therapy; gene; single nucleotide polymorphism; SNP;
KW      chromosome 12p13-p12.3; 56.
XX
OS      Homo sapiens.
XX
FH      Key
FH      Location/Qualifiers
FT      variation
FT      replace(112,a)
FT      /tag= a
FT      /label= RAI-3-s1
FT      /note= "located in 5' untranslated region"
FT      /standard_name= "Single nucleotide polymorphism"
FT      CDS
FT      254..1327
FT      /tag= b
FT      /product= "Human RAI3"
FT      replace(364,t)
FT      /tag= c
FT      /label= RAI-3-s2
FT      /note= "exon 1, silent (Ala/Ala)"
FT      /standard_name= "Single nucleotide polymorphism"
FT      variation
FT      replace(511,t)
FT      /tag= d
FT      /label= RAI-3-s3
FT      /note= "exon 2, silent (Ile/Ile)"
FT      /standard_name= "Single nucleotide polymorphism"
FT      variation
FT      replace(523,t)
FT      /tag= e
FT      /label= RAI-3-s4
FT      /note= "exon 2, silent (Asp/Asp)"
FT      /standard_name= "Single nucleotide polymorphism"
FT      variation
FT      replace(605,g)
FT      /tag= f
FT      /label= RAI-3-s6
FT      /note= "exon 2, missense (Ser/Gly)"
FT      /standard_name= "Single nucleotide polymorphism"
FT      variation
FT      replace(797,g)
FT      /tag= g
FT      /label= RAI-3-s5
FT      /note= "exon 2, missense (Thr/Ala)"
FT      /standard_name= "Single nucleotide polymorphism"
FT      variation
FT      replace(1111,g)
FT      /tag= h
FT      /label= RAI-3-s8
FT      /note= "silent (Pro/Pro)"
FT      /standard_name= "Single nucleotide polymorphism"
FT      variation
FT      replace(1173,g)
FT      /tag= i
FT      /label= RAI-3-s9
FT      /note= "missense (Gln/Arg)"
FT      /standard_name= "Single nucleotide polymorphism"

```

```

XX      WO2004001060-A2.
XX
XX      31-DEC-2003.
XX
XX      20-JUN-2003; 2003WO-US019255.
XX
XX      20-JUN-2002; 2002US-0390850P.
XX      29-AUG-2002; 2002US-0407006P.
XX
XX      (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
XX      Whitney GS, Opitck G, Garulacan L, Ramanathan CS, McKinnon M;
XX      Bennett KL, Barber LE, Cacace A, Teuchihashi Z;
XX
XX      WPI; 2004-090973/09.
XX      P-PSDB; ADI28460.
XX      GENBANK; NM_003979.
XX
XX      New nucleic acid molecule encoding a human G-protein coupled receptor,
XX      RAI3, useful for preventing, treating or ameliorating chronic obstructive
XX      pulmonary disease (COPD), COPD-like disorder, or the underlying symptoms
XX      of COPD.
XX
XX      Claim 1; SEQ ID NO 2; 301pp; English.
XX
XX      The present sequence is that of cDNA encoding a human G-protein coupled
XX      receptor (Class C, Group 5), denoted retinoic acid induced 3 (RAI-3).
XX      Proteomics methods were used to isolate cigarette smoke-inducible
XX      tyrosine phosphorylated proteins from airway epithelial cells. RAI-3 was
XX      identified as being tyrosine phosphorylated and/or as being
XX      associated/complexed with tyrosine phosphorylated proteins only in those
XX      cells that had been exposed to cigarette smoke. Since RAI-3 is primarily
XX      expressed in lung tissue, and since cigarette smoke is a major causative
XX      factor of chronic obstructive pulmonary disease (COPD), RAI-3 provides a
XX      novel cellular target for identifying modulators, e.g. agonists or
XX      antagonists, useful for the treatment and/or prevention of COPD and
XX      related disorders such as emphysema and chronic bronchitis. RAI-3
XX      modulators, e.g. agonists and antagonists, especially antisense
XX      compounds, can be used to treat COPD and other disorders and diseases
XX      associated with regulation of NF-kB and/or its associated or interacting
XX      signaling molecules. Single nucleotide polymorphisms (SNPs) detected in
XX      the RAI-3 gene are useful for determining COPD association in
XX      individuals. RAI-3 nucleic acid molecules and polypeptides are useful for
XX      preventing, treating or ameliorating disorders related to aberrant GPCR
XX      signaling or cell cycle regulation, pulmonary disorders, inflammatory
XX      lung disorders, COPD, the underlying symptoms of COPD, COPD-related
XX      disorders or conditions, autoimmune disorders, disorders related to
XX      hyperimmune activity, inflammatory conditions, disorders related to
XX      aberrant acute phase responses, hypercongenital conditions, birth
XX      defects, necrotic lesions, wounds, organ transplant rejection, renal
XX      diseases, ischaemia-reperfusion injury, heart disorders, disorders
XX      related to aberrant signal transduction, proliferation disorders,
XX      cancers, HIV infection, asthma, cystic fibrosis, pulmonary fibrosis,
XX      ulcerative colitis, cerebral infarct, myocardial infarct, diabetic
XX      nephropathy, allergic rhinitis, Crohn's disease, atherosclerosis,
XX      rheumatoid arthritis, inflammatory/autoimmune disorders, glioblastoma,
XX      pulmonary small cell undifferentiated carcinoma, carcinoma of the breast,
XX      colon, lung, ovary, pancreas, prostate, non-Hodgkin's lymphoma, disorders
XX      associated with aberrant cell adhesion, I-CAM function and/or regulation,
XX      B-selectin function and/or regulation, or aberrant NF-kB function and/or
XX      regulation (all claimed).
XX
XX      Sequence 2456 BP; 519 A; 722 C; 571 G; 644 T; 0 U; 0 Other;
XX
XX      Query Match      100.0%; Score 2456; DB 12; Length 2456;
XX      Best Local Similarity 100.0%; Pred. No. 0;
XX      Matches 2456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1  ATAAAGCATGAAGTCCGTTGGAACTGGAAATAGCGGTGCTCTCTCCCTGACCTCCCCC 60
Db      1  ATAAAGCATGAAGTCCGTTGGAACTGGAAATAGCGGTGCTCTCTCCCTGACCTCCCCC 60

```

QY 61 TCCTTGTCCCTCTGCTCAGCCCTCGCTCGTTCCTCCCTCCCTCCGCGAGGGCCGCTTTATA 120  
Db |||||  
QY 61 TCCTTGTCCCTCTGCTCAGCCCTCGCTCGTTCCTCCCTCCGCGAGGGCCGCTTTATA 120  
Db |||||  
QY 121 ACAACTGCTCAGAGTGCAGAGGGCGGATAGCTGTCCAAGGTCTCCGCCAGCACTAGAGGAG 180  
Db |||||  
QY 121 ACAACTGCTCAGAGTGCAGAGGGCGGATAGCTGTCCAAGGTCTCCGCCAGCACTAGAGGAG 180  
Db |||||  
QY 181 CTCGCCGTGTCCTCTTCGCGCGCGGAAGAGCAGCACAAAGTTCA CGGCCAAACGCTTTGCC 240  
Db |||||  
QY 181 CTCGCCGTGTCCTCTTCGCGCGCGGAAGAGCAGCACAAAGTTCA CGGCCAAACGCTTTGCC 240  
Db |||||  
QY 241 ACTAGGGTCCAGATGGCTACACAGTCCCTGATGGTTGCGGCAATGGCTGAAATCCAA 300  
Db |||||  
QY 241 ACTAGGGTCCAGATGGCTACACAGTCCCTGATGGTTGCGGCAATGGCTGAAATCCAA 300  
Db |||||  
QY 301 GTACTACAGACTTTGTGATAAGGCTGAAGCTTGGGGCATCGCTTAGAAAACGGTGGCCAC 360  
Db |||||  
QY 301 GTACTACAGACTTTGTGATAAGGCTGAAGCTTGGGGCATCGCTTAGAAAACGGTGGCCAC 360  
Db |||||  
QY 361 AGCCGGGTTGTGACCTCGGTGGCTTCATGTCTACTCTCCGATCTCGTCTGCAAGGT 420  
Db |||||  
QY 361 AGCCGGGTTGTGACCTCGGTGGCTTCATGTCTACTCTCCGATCTCGTCTGCAAGGT 420  
Db |||||  
QY 421 GCAGGACTCCAAAGCGGAAATGCTGCTACTCAGTTTCTCTTCTCTCTCTGGGTGTGT 480  
Db |||||  
QY 421 GCAGGACTCCAAAGCGGAAATGCTGCTACTCAGTTTCTCTTCTCTCTCTGGGTGTGT 480  
Db |||||  
QY 481 GGGCATCTTTGGCTCACCTTCGCCCTTCATCATCGGACTGGAGCGGAGCACAGGGCCAC 540  
Db |||||  
QY 481 GGGCATCTTTGGCTCACCTTCGCCCTTCATCATCGGACTGGAGCGGAGCACAGGGCCAC 540  
Db |||||  
QY 541 AGCTTCTTCTCTTTGGGATCCTCTTTTCCATGCTCTCTCTCGCTCGCTGGCTCATGC 600  
Db |||||  
QY 541 AGCTTCTTCTCTTTGGGATCCTCTTTTCCATGCTCTCTCTCGCTCGCTGGCTCATGC 600  
Db |||||  
QY 601 TGTGAGTCTGACCAAGCTCGTCCGGGGAGGAGCCCTTCCCTGTGGTGAATCTGGG 660  
Db |||||  
QY 601 TGTGAGTCTGACCAAGCTCGTCCGGGGAGGAGCCCTTCCCTGTGGTGAATCTGGG 660  
Db |||||  
QY 661 TCTGCGCTGTGGCTTCAGCCTAGTCAGGATGTTATCGCTATTGAATATATTGTCCTGAC 720  
Db |||||  
QY 661 TCTGCGCTGTGGCTTCAGCCTAGTCAGGATGTTATCGCTATTGAATATATTGTCCTGAC 720  
Db |||||  
QY 721 CATGAATAGGACCAACGTCATGCTTTTCTGAGCTTCCGCTCTCTGTCGCAATGAAGA 780  
Db |||||  
QY 721 CATGAATAGGACCAACGTCATGCTTTCTGAGCTTCCGCTCTCTGTCGCAATGAAGA 780  
Db |||||  
QY 781 CTTTGTCTCTCTGCTCAGCTAGCTCTCTCTTCTGATGGGCTGACCTTCTCATGTCCTC 840  
Db |||||  
QY 781 CTTTGTCTCTCTGCTCAGCTAGCTCTCTCTTCTGATGGGCTGACCTTCTCATGTCCTC 840  
Db |||||  
QY 841 CTTTCACTTCTGTGGTTCTTCA CGGGCTGGAAGAGACATGGGGCCACATCTACTCTAC 900  
Db |||||  
QY 841 CTTTCACTTCTGTGGTTCTTCA CGGGCTGGAAGAGACATGGGGCCACATCTACTCTAC 900  
Db |||||  
QY 901 GATGTCCTCTCCATTGCGCATCTGGGTGGCTGGATCACCTGCTCATGCTTCTCTGACTT 960  
Db |||||  
QY 901 GATGTCCTCTCCATTGCGCATCTGGGTGGCTGGATCACCTGCTCATGCTTCTCTGACTT 960  
Db |||||  
QY 961 TGACCGCAGGTGGGATGACACCATCTCAGCTCCGCTTGGCTGGCTGCAATGGCTGGGTGT 1020  
Db |||||  
QY 961 TGACCGCAGGTGGGATGACACCATCTCAGCTCCGCTTGGCTGGCTGCAATGGCTGGGTGT 1020  
Db |||||  
QY 1021 CCTGTTGGCTTATGTTAGTCCCGAGTTTGGCTGCTCAGAAAGCAACGAAACCCCATGGA 1080  
Db |||||  
QY 1021 CCTGTTGGCTTATGTTAGTCCCGAGTTTGGCTGCTCAGAAAGCAACGAAACCCCATGGA 1080  
Db |||||  
QY 1081 TTATCCTGTTAGGATGCTTTCTGTAAACCTCAACTCTGTGGAAGAGACTATGGTGTGA 1140  
Db |||||  
QY 1081 TTATCCTGTTAGGATGCTTTCTGTAAACCTCAACTCTGTGGAAGAGACTATGGTGTGA 1140  
Db |||||  
QY 1141 GNAACAGCCTACTCTCAAGAGGAAATCACTCAAGGTTTGAAGAGACAGGGGACACGCT 1200

Db |||||  
QY 1141 GAACAGAGCCTACTCTCAAGAGGAAATCACTCAAGGTTTGAAGACAGGGGACACGCT 1200  
Db |||||  
QY 1201 CTATGCCCCCTATTCCACACATTTTTCAGCTGCAGAACACAGCCTCCCCAAAAGGAAATCTC 1260  
Db |||||  
QY 1201 CTATGCCCCCTATTCCACACATTTTTCAGCTGCAGAACACAGCCTCCCCAAAAGGAAATCTC 1260  
Db |||||  
QY 1261 CATCCACAGGGGCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAGAAAGAGGG 1320  
Db |||||  
QY 1261 CATCCACAGGGGCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAGAAAGAGGG 1320  
Db |||||  
QY 1321 CAGCTAACTCTGCTCTGAGAGTGGGCAAAATGCGACCGGGCGGAGACTAGCGGAGC 1380  
Db |||||  
QY 1321 CAGCTAACTCTGCTCTGAGAGTGGGCAAAATGCGACCGGGCGGAGACTAGCGGAGC 1380  
Db |||||  
QY 1381 TCAAAAGGAGTGGGCGAAATCTTGAGTCTTCTGAGAAAATGTA CAAGACACTACGGGA 1440  
Db |||||  
QY 1381 TCAAAAGGAGTGGGCGAAATCTTGAGTCTTCTGAGAAAATGTA CAAGACACTACGGGA 1440  
Db |||||  
QY 1441 ACAGTTTGTCTCCCTCCAGCCTCAACCAAAATTTCTTCATGCTGGGGCTGATGGGCT 1500  
Db |||||  
QY 1441 ACAGTTTGTCTCCCTCCAGCCTCAACCAAAATTTCTTCATGCTGGGGCTGATGGGCT 1500  
Db |||||  
QY 1501 AGTAAGACTCCAGTTCTTAGAGGCGCTGTAGTATTTTCTTTTCTCTCATCTCTTTGG 1560  
Db |||||  
QY 1501 AGTAAGACTCCAGTTCTTAGAGGCGCTGTAGTATTTTCTTTTCTCTCATCTCTTTGG 1560  
Db |||||  
QY 1561 ATACTTCTTTTAAAGTGGGAGTCTCAGGCAACTCAAGTTTGTAGACCCCTTACTCTTTTGT 1620  
Db |||||  
QY 1561 ATACTTCTTTTAAAGTGGGAGTCTCAGGCAACTCAAGTTTGTAGACCCCTTACTCTTTTGT 1620  
Db |||||  
QY 1621 GTTTTTTGAAACAGAGATCTTGTCTGTCAACCAGGCTTGAGTGCAAGTGTGATCACAG 1680  
Db |||||  
QY 1621 GTTTTTTGAAACAGAGATCTTGTCTGTCAACCAGGCTTGAGTGCAAGTGTGATCACAG 1680  
Db |||||  
QY 1681 CCCAGTGCAGCTCCAGCACCTGTCTCAAGCAATCTCCCATCTCCATCTCCCAAAGTG 1740  
Db |||||  
QY 1681 CCCAGTGCAGCTCCAGCACCTGTCTCAAGCAATCTCCCATCTCCATCTCCCAAAGTG 1740  
Db |||||  
QY 1741 CTGGGATGACAGCGCTGAGCCACAGCTCCAGCCTAGGCCCTTAATCTTGTCTGTTATTTT 1800  
Db |||||  
QY 1741 CTGGGATGACAGCGCTGAGCCACAGCTCCAGCCTAGGCCCTTAATCTTGTCTGTTATTTT 1800  
Db |||||  
QY 1801 COATGGACTAAAGTCTGGTCACTGAGCTCAGCTGGCTCAACAGCTCTTAGGGGCTG 1860  
Db |||||  
QY 1801 COATGGACTAAAGTCTGGTCACTGAGCTCAGCTGGCTCAACAGCTCTTAGGGGCTG 1860  
Db |||||  
QY 1861 CTCCTCTAATCACAAGTGGGTTTGTGAGGCTCTGTGGCCAGAGCAGACCTGCATATCT 1920  
Db |||||  
QY 1861 CTCCTCTAATCACAAGTGGGTTTGTGAGGCTCTGTGGCCAGAGCAGACCTGCATATCT 1920  
Db |||||  
QY 1921 GAGCAAAAATAGCAAAAGCCTCTCTCAGCCCACTGGCCCTGAATCTACACTGGAAGCCAA 1980  
Db |||||  
QY 1921 GAGCAAAAATAGCAAAAGCCTCTCTCAGCCCACTGGCCCTGAATCTACACTGGAAGCCAA 1980  
Db |||||  
QY 1981 TTGCTGGGACCCCGCTCCCAAACCTTCTTGGCTGGGTAGGAGGCTTAAAGATCACCC 2040  
Db |||||  
QY 1981 TTGCTGGGACCCCGCTCCCAAACCTTCTTGGCTGGGTAGGAGGCTTAAAGATCACCC 2040  
Db |||||  
QY 2041 TAAATTTTACTCATCTCTAGTGTGCTCAATTTGGGCTCAGAGCTCCCGAGCACA 2100  
Db |||||  
QY 2041 TAAATTTTACTCATCTCTAGTGTGCTCAATTTGGGCTCAGAGCTCCCGAGCACA 2100  
Db |||||  
QY 2101 ATTCAAGGTCAACCCCTCTCTTTTGCATGTGCCCAAACTTGTCTCAATTTCCGAGATC 2160  
Db |||||  
QY 2161 TAAATCTCCCTACGCTCTGCCAGAAATCTTTTCAAGCTCTACTAGCACAAAGCCGGTTG 2220  
Db |||||  
QY 2161 TAAATCTCCCTACGCTCTGCCAGAAATCTTTTCAAGCTCTACTAGCACAAAGCCGGTTG 2220  
Db |||||  
QY 2221 CTCCTTGTGAGGAGAAATTTGTAGATCATTTCTCAATTTCAAATTTCTCTGGGCTGATCTCT 2280  
Db |||||

Db 2221 CTCCTTGTCCAGGAAATTTGTAGATCATTTCTCAATCTCTCAAAATCTCTGGGCTGATCTTCT 2280  
Qy 2281 CTATCTTGCACCCCAACCTCTGTAATAGATTTACCGCATTTACGGCTGCTGCTGTA 2340  
Db 2281 TTTTCTTGTGACCCCAACCTCTGTAATAGATTTACCGCATTTACGGCTGCTGCTGTA 2340  
Qy 2341 GTGGGATGCTCTCCTAATAGGAGAGTGTTCATTTGATATATAAGTTATTACCTGAGTAT 2400  
Db 2341 GTGGGATGCTCTCCTAATAGGAGAGTGTTCATTTGATATATAAGTTATTACCTGAGTAT 2400  
Qy 2401 GCAATAAGATGTGGTGGCCACTCTTTTCATGGTGGTGGCAGCAAAAAA 2456  
Db 2401 GCAATAAGATGTGGTGGCCACTCTTTTCATGGTGGTGGCAGCAAAAAA 2456

## RESULT 4

ADR48221  
ID ADR48221 standard; cDNA; 2456 BP.

XX AC ADR48221;

XX DT 18-NOV-2004 (first entry)

XX DE Human retinoic acid induced 3 (RAI3) encoding cDNA SEQ.9.

XX KW pancreatic cancer-associated transcript; pancreatic cancer; human;  
KW cytostatic; gene therapy; protein therapy; retinoic acid induced 3; RAI3;  
KW gene; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers  
FT CDS 254..1327

FT FT /tag= a  
FT FT /product= "retinoic acid induced 3 (RAI3)"

XX PN WO2004074510-A1.

XX PD 02-SEP-2004.

XX PF 18-FEB-2004; 2004WO-AU000194.

XX PR 18-FEB-2003; 2003AU-00900747.

XX PA (GARV-) GARVAN INST MEDICAL RES.

XX PI Biankin A, Segara D, Henshall S, Sutherland R;

XX DR WPI; 2004-635591/61.

XX DR P-PSDB; ADR48222.

PT Detecting pancreatic cancer-associated transcript in a biological sample,  
PT useful for diagnosing or treating the disease, comprises contacting the  
PT sample with a polynucleotide that selectively hybridizes to a specific  
PT sequence.

XX PS Claim 7; SEQ ID NO 9; 263pp; English.

XX CC The present invention describes a method for detecting a pancreatic  
CC cancer-associated transcript in a biological sample. The method comprises  
CC contacting the biological sample with a polynucleotide that selectively  
CC hybridizes to a sequence at least 80% identical to a sequence as shown in  
CC any one of Tables 3 to 25 in the specification or having the GenBank  
CC Accession Number AF29145. Also described: (1) diagnosing pancreatic  
CC cancer in a human or animal subject being tested, determining the  
CC likelihood that a subject having a pancreatic cancer will survive, or  
CC determining the suitability of a subject having a pancreatic cancer for  
CC surgical resection therapy; (2) detecting a pancreatic cancer-associated  
CC polypeptide in a biological sample; (3) determining the likelihood that a  
CC subject having a pancreatic cancer will survive; and (4) monitoring the  
CC efficacy of a therapeutic treatment of pancreatic cancer. A pancreatic  
CC cancer-associated transcript has cytostatic activity, and can be used in  
CC gene and protein therapy. A pancreatic cancer-associated transcript

CC polynucleotide, a vector comprising the polynucleotide, an isolated  
CC polypeptide or an antibody that binds to the isolated polypeptide can be  
CC used for diagnosing or prognosing pancreatic cancer or for preparing a  
CC medicament for the treatment of pancreatic cancer. The prognostic or  
CC diagnostic methods are useful for the early detection of pancreatic  
CC cancer or its metastases, and for monitoring the progress of disease such  
CC as during remission or following surgery or chemotherapy. The present  
CC sequence encodes human retinoic acid induced 3 (RAI3), which is used in  
CC the exemplification of the present invention.

XX SQ Sequence 2456 BP; 519 A; 722 C; 571 G; 644 T; 0 U; 0 Other;

Query Match 100.0%; Score 2456; DB 13; Length 2456;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATAACAGCATGAAGTCCGCTGGAACTGGAAATAGGCGTGTCTCTCCCTCGAGGGCGCTTTATA 120

Db 1 ATAACAGCATGAAGTCCGCTGGAACTGGAAATAGGCGTGTCTCTCCCTCGAGGGCGCTTTATA 120

Qy 61 TCCTTGTCCCTCTGCTCAGCCCTCGCTCGTTCCTCCCTCCGCGAGGGCGCTTTATA 120

Db 61 TCCTTGTCCCTCTGCTCAGCCCTCGCTCGTTCCTCCCTCCGCGAGGGCGCTTTATA 120

Qy 121 ACAACTGCTCAGAGTGGAGGGCGGATAGCTGTCCAAGGTCTCCCCAGCACTGAGGAG 180

Db 121 ACAACTGCTCAGAGTGGAGGGCGGATAGCTGTCCAAGGTCTCCCCAGCACTGAGGAG 180

Qy 181 CTGCGCTGTGCCCTCTTGGCGCGGGAAGCAGCACAAGTTCCACGGCCAAACGCTTGGC 240

Db 181 CTGCGCTGTGCCCTCTTGGCGCGGGAAGCAGCACAAGTTCCACGGCCAAACGCTTGGC 240

Qy 241 ACTAGGGTCCAGATGGCTACACAGTCCCTGATGTTCCGCAATGGCTGAAATCCAA 300

Db 241 ACTAGGGTCCAGATGGCTACACAGTCCCTGATGTTCCGCAATGGCTGAAATCCAA 300

Qy 301 GTACTACAGACTTTGTGATAAGGCTGAAGCTTGGGGCATCGTCTAGAAACGGTGGCCAC 360

Db 301 GTACTACAGACTTTGTGATAAGGCTGAAGCTTGGGGCATCGTCTAGAAACGGTGGCCAC 360

Qy 361 AGCCGGGGTTGTGACCTCGGTGGCTTCATGCTCACTCTCCGATCCTCGTCTGCAAGGT 420

Db 361 AGCCGGGGTTGTGACCTCGGTGGCTTCATGCTCACTCTCCGATCCTCGTCTGCAAGGT 420

Qy 421 GCAGGACTCCAAACAGCGGAAATGCTGCTACTCAGTTTCTCTTCTCTGGGTGTGT 480

Db 421 GCAGGACTCCAAACAGCGGAAATGCTGCTACTCAGTTTCTCTTCTCTGGGTGTGT 480

Qy 481 GGGCATCTTTGGCTCACCCTTCGCTTCATCATCGGACTGGAGCGGAGCAGCGCCAC 540

Db 481 GGGCATCTTTGGCTCACCCTTCGCTTCATCATCGGACTGGAGCGGAGCAGCGCCAC 540

Qy 541 AGCTTCTTCTCTTTGGGATCCTCTTTTCCATCTGCTTCTCTGCTGCTGCTCATGC 600

Db 541 AGCTTCTTCTCTTTGGGATCCTCTTTTCCATCTGCTTCTCTGCTGCTGCTCATGC 600

Qy 601 TGTCTAGTCTGACCAAGCTCGTCCGGGGAGGAGCCCTTCCCTCTGCTGATTCCTGG 660

Db 601 TGTCTAGTCTGACCAAGCTCGTCCGGGGAGGAGCCCTTCCCTCTGCTGATTCCTGG 660

Qy 661 TCTGGCCGTGGGCTTCAGCCTAGTCCAGGATGTTATCGCTATTGAAATATATATCTCTGAC 720

Db 661 TCTGGCCGTGGGCTTCAGCCTAGTCCAGGATGTTATCGCTATTGAAATATATATCTCTGAC 720

Qy 721 CATGAATAGGACCAACAGTCAATGTCTTTTCTGAGCTTTCGGTCTCTCGTGGCAATGAAGA 780

Db 721 CATGAATAGGACCAACAGTCAATGTCTTTTCTGAGCTTTCGGTCTCTCGTGGCAATGAAGA 780

Qy 781 CTTTGTCTCTCTGCTCAGCTAGTCTCTCTTCTTGTAGGGCTGACCTTCTCATGTCCTC 840

Db 781 CTTTGTCTCTCTGCTCAGCTAGTCTCTCTTCTTGTAGGGCTGACCTTCTCATGTCCTC 840

Qy 841 CTTTACCTTCTGTGGTTCCTTTCACGGGCTGGAAGAGACATGGGGGCCACATCTACCTCAC 900

Db 841 CTTACCTTTCTGGTTCCTTACGGGCTGGAGACATGGGGCCACATCTACCTCAC 900  
Qy 901 GATGTCCTCTCATTGCCATCTGGGTGGCTGGATCACCGTCTCATCTCTCTGACTT 960  
Db 901 GATGTCCTCTCATTGCCATCTGGGTGGCTGGATCACCGTCTCATCTCTCTGACTT 960  
Qy 961 TGACCGCAGGTGGGATGACACCATCTCAGCTCCGCTTTGGCTGGCOAATGGCTGGGTGT 1020  
Db 961 TGACCGCAGGTGGGATGACACCATCTCAGCTCCGCTTTGGCTGGCOAATGGCTGGGTGT 1020  
Qy 1021 CTGTTGGCTTATGTTAGTCCGAGTTTGGCTGGCTCAAAAGCAACGAAACCCCATGGA 1080  
Db 1021 CTGTTGGCTTATGTTAGTCCGAGTTTGGCTGGCTCAAAAGCAACGAAACCCCATGGA 1080  
Qy 1081 TTATCTGTTGAGGATGCTTTCTGTAACCTCAACTCGTGAAGAGAGCTATGGTGTGGA 1140  
Db 1081 TTATCTGTTGAGGATGCTTTCTGTAACCTCAACTCGTGAAGAGAGCTATGGTGTGGA 1140  
Qy 1141 GAACGAGCCTACTCTCAAGAGGAATCACTCAAGGTTTTGAAGAGACAGGGACACGCT 1200  
Db 1141 GAACGAGCCTACTCTCAAGAGGAATCACTCAAGGTTTTGAAGAGACAGGGACACGCT 1200  
Qy 1201 CTATGCCCTTATTCACACATTTTCAGCTGCAGACACGCTCCCCAAAGGAATTC 1260  
Db 1201 CTATGCCCTTATTCACACATTTTCAGCTGCAGACACGCTCCCCAAAGGAATTC 1260  
Qy 1261 CATCCACGGGCCACGCTTGGCCGAGCCTTACAAAGACTATGAAGTAAAGAAAGAGG 1320  
Db 1261 CATCCACGGGCCACGCTTGGCCGAGCCTTACAAAGACTATGAAGTAAAGAAAGAGG 1320  
Qy 1321 CAGCTAACTCTGTCTGAAGAGTGGACAAATGCAGCCGGCGGCGAGATCTAGCGGGAGC 1380  
Db 1321 CAGCTAACTCTGTCTGAAGAGTGGACAAATGCAGCCGGCGGCGAGATCTAGCGGGAGC 1380  
Qy 1381 TCNAAGGATGGGGGGAATCTTGATCTCTGAGAAAACCTGTACAGACACTACGGGA 1440  
Db 1381 TCNAAGGATGGGGGGAATCTTGATCTCTGAGAAAACCTGTACAGACACTACGGGA 1440  
Qy 1441 ACAGTTTGCTCCTCCAGCCTCAACACAAATTTCTCATGCTGGGGCTGATGGGCT 1500  
Db 1441 ACAGTTTGCTCCTCCAGCCTCAACACAAATTTCTCATGCTGGGGCTGATGGGCT 1500  
Qy 1501 AGTAAGCTCCAGTCTTGAAGCGCTGTAGATATTTTTTTTTTTTGTCTCATCTCTGG 1560  
Db 1501 AGTAAGCTCCAGTCTTGAAGCGCTGTAGATATTTTTTTTTTTTGTCTCATCTCTGG 1560  
Qy 1561 ATACTTCTTTTAAAGTGGAGTCTCAGGCACTCAAGTTTAGACCTTACTCTTTTGT 1620  
Db 1561 ATACTTCTTTTAAAGTGGAGTCTCAGGCACTCAAGTTTAGACCTTACTCTTTTGT 1620  
Qy 1621 GTTTTTTGAACAGGATCTTGCTCTGTCAACCCAGGCTTGAGTGCAGTGGCGATCACAG 1680  
Db 1621 GTTTTTTGAACAGGATCTTGCTCTGTCAACCCAGGCTTGAGTGCAGTGGCGATCACAG 1680  
Qy 1681 CCAGTGCAGCTTCGACACCTGTGTCTCAAGCAATCTCCCATCTCCATCTCCCAAAGTG 1740  
Db 1681 CCAGTGCAGCTTCGACACCTGTGTCTCAAGCAATCTCCCATCTCCATCTCCCAAAGTG 1740  
Qy 1741 CTGGGATGACAGCGTGAAGCAGCTCCAGCCTAGGCGCTTAATCTTGCTGTATTTT 1800  
Db 1741 CTGGGATGACAGCGTGAAGCAGCTCCAGCCTAGGCGCTTAATCTTGCTGTATTTT 1800  
Qy 1801 CCATGACTAAAGGTCTGTCTCATCTCAGCTCAGCTGGCTCACACAGCTCTAGGGCCTG 1860  
Db 1801 CCATGACTAAAGGTCTGTCTCATCTCAGCTCAGCTGGCTCACACAGCTCTAGGGCCTG 1860  
Qy 1861 CTCCTCTAACTCAAGTGGTTTTGTAGGCTCTGTGGCCAGAGCAGACCTTGCATATCT 1920  
Db 1861 CTCCTCTAACTCAAGTGGTTTTGTAGGCTCTGTGGCCAGAGCAGACCTTGCATATCT 1920  
Qy 1921 GAGCAAAATAGCAAAAGCTCTCTCAGCCCACTGGCCCTGAATCTACACTGGAAGCAAC 1980

Db 1921 GAGCAAAATAGCAAAAGCTCTCTCAGCCCACTGGCCTGAATCTACACTGGAGCAAC 1980  
Qy 1981 TTGCTGGCACCCCGCTCCCAACCCCTTCTTGCTGGGTAGGAGGGCTAAAGATCACCC 2040  
Db 1981 TTGCTGGCACCCCGCTCCCAACCCCTTCTTGCTGGGTAGGAGGGCTAAAGATCACCC 2040  
Qy 2041 TAAATTTACTCATCTCTTAGTGTGCTCATATGGGCTCAGCAGCTCCCCAGACCA 2100  
Db 2041 TAAATTTACTCATCTCTTAGTGTGCTCATATGGGCTCAGCAGCTCCCCAGACCA 2100  
Qy 2101 ATTACAGGTACCCCTCTCTTTTGCACTGCCCAACTTGTCTCAATTCGAGATC 2160  
Db 2101 ATTACAGGTACCCCTCTCTTTTGCACTGCCCAACTTGTCTCAATTCGAGATC 2160  
Qy 2161 TAATCTCCCTTACGCTCTGCCAGGAATTTCTTTCAGACCTCACTAGCAAAAGCCCGTTG 2220  
Db 2161 TAATCTCCCTTACGCTCTGCCAGGAATTTCTTTCAGACCTCACTAGCAAAAGCCCGTTG 2220  
Qy 2221 CTCCTTGTACAGGAATTTGTAGATCATTTCACTTCAATTCCTGGGCTGATCTTCT 2280  
Db 2221 CTCCTTGTACAGGAATTTGTAGATCATTTCACTTCAATTCCTGGGCTGATCTTCT 2280  
Qy 2281 CTCATCTTCCACCCCAACCTCTGTAATAGATTTACCGCATTTACGGCTGCATTTCTGTA 2340  
Db 2281 CTCATCTTCCACCCCAACCTCTGTAATAGATTTACCGCATTTACGGCTGCATTTCTGTA 2340  
Qy 2341 GTGGCATGGTCTCCTTAATGGAGAGTGTTCATTTGATTAATAGTTTATTCACCTGAGTAT 2400  
Db 2341 GTGGCATGGTCTCCTTAATGGAGAGTGTTCATTTGATTAATAGTTTATTCACCTGAGTAT 2400  
Qy 2401 GCAATAAAGATGTGGTGGCCTCTTTTCATGTTGGTGGCAGCAAAAAGAAAAA 2456  
Db 2401 GCAATAAAGATGTGGTGGCCTCTTTTCATGTTGGTGGCAGCAAAAAGAAAAA 2456

## RESULT 5

ACN39388  
ID ACN39388 standard; cDNA; 2456 BP.

AC ACN39388;

DT 18-NOV-2004 (first entry)

XX Tumour-associated antigenic target (TAT) cDNA DNA226771, SEQ ID NO:3498.

DE Tumour-associated antigenic target; TAT; human; overexpression; cancer;

KW tumour; diagnosis; cell proliferative disorder; breast cancer;

KW colorectal cancer; lung cancer; ovarian cancer; liver cancer;

KW central nervous system cancer; bladder cancer; pancreatic cancer;

KW cervical cancer; melanoma; leukaemia; hybridisation probe;

KW chromosome identification; chromosome mapping; gene mapping;

XX Homo sapiens.

OS Homo sapiens.

XX WO2004030615-A2.

XX 15-APR-2004.

PD 29-SEP-2003; 2003WO-US028547.

PP 02-OCT-2002; 2002US-0414971P.

XX (GETH ) GENENTECH INC.

XX Wu TD, Zhang Z, Zhou Y;

XX WPI; 2004-347921/32.

DR P-PSDB; ARB81354.

XX New tumor-associated antigenic target polypeptides and nucleic acids, useful in preparing a medicament for treating or detecting a proliferative disorder, e.g. breast, lung, colorectal, ovarian or

prostate cancer or tumor.

Claim 1; SEQ ID NO 3498; 7273pp; English.

The invention relates to human tumour-associated, antigenic target (TAT) polypeptides, and their related nucleic acids. The TAT polypeptides are overexpressed in cancer tissues compared to normal tissues, and may thus serve as effective targets for the diagnosis and treatment of cancer in mammals. The invention also relates to nucleic acid and polypeptide sequences at least 80% identical to the TAT nucleic acids and polypeptides; expression vectors and host cells comprising a TAT nucleic acid; an antibody specific for a TAT polypeptide; a peptide or organic molecule which binds to a TAT polypeptide; fusion proteins comprising a TAT polypeptide; and methods and compositions for the treatment or diagnosis of cancer in mammals. TAT polypeptides, nucleic acids, antibodies, antagonists, binding molecules and compositions are useful for diagnosing or treating a cell proliferative disorder associated with increased TAT expression, particularly cancers such as breast cancer, colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder cancer, pancreatic cancer, cervical cancer, cancers of the central nervous system, melanoma and leukaemia. TAT nucleic acids may further be used as hybridisation probes, in chromosome and gene mapping, in chromosome identification and in gene therapy. The present sequence represents a TAT nucleic acid of the invention.

Sequence 2456 BP; 519 A; 722 C; 571 G; 644 T; 0 U; 0 Other;

Query Match 100.0%; Score 2456; DB 13; Length 2456;

Best Local Similarity

Matches 2456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	ATAACAGCATGAAGTCGCGTGAACCTGGAAATAGGCGGTGTCCTCTCCCTCGAACCTCTCCCC	60
Db	1	ATAACAGCATGAAGTCGCGTGAACCTGGAAATAGGCGGTGTCCTCTCCCTCGAACCTCTCCCC	60
Qy	61	TCCTTGTCCTCTGCTCAACCCCTCGCTCGTTCCTCTCCCTCCGGCAGGCGCGCTTTATA	120
Db	61	TCCTTGTCCTCTGCTCAACCCCTCGCTCGTTCCTCTCCCTCCGGCAGGCGCGCTTTATA	120
Qy	121	ACAACCTGCTCAGAGTCGAGGGCGGGATAGCTGTCAAGGTCTCCCCAGCACTGAGGAG	180
Db	121	ACAACCTGCTCAGAGTCGAGGGCGGGATAGCTGTCAAGGTCTCCCCAGCACTGAGGAG	180
Qy	181	CTCGCTCTGCTGCCCTCTTTGCGCGCGGGAAGCAGCACCAAGTTTCAGGGCCAAACGCTTTGGC	240
Db	181	CTCGCTCTGCTGCCCTCTTTGCGCGCGGGAAGCAGCACCAAGTTTCAGGGCCAAACGCTTTGGC	240
Qy	241	ACTAGGGTTCAGAAATGGCTACAAAGTCCCTGATGGTTGCGCAATGGCTGAAATCCAA	300
Db	241	ACTAGGGTTCAGAAATGGCTACAAAGTCCCTGATGGTTGCGCAATGGCTGAAATCCAA	300
Qy	301	GTACTACAGACTTTGTGTGATAAGGCTGAAGCTTGGGGCATCGTCTAGAAAACGGTGGCCAC	360
Db	301	GTACTACAGACTTTGTGTATAGGCTGAAGCTTGGGGCATCGTCTAGAAAACGGTGGCCAC	360
Qy	361	AGCCGGGGTTGTGACCTCGGTGGCCCTTCATGCTCACTCTCCGATCCTCGTCTGCAAGGT	420
Db	361	AGCCGGGGTTGTGACCTCGGTGGCCCTTCATGCTCACTCTCCGATCCTCGTCTGCAAGGT	420
Qy	421	GCAGGACTCCAAACAGGCGGAAAAATGCTGCTACTCAGTTTTCTCTCTCCTCGGTGTGTT	480
Db	421	GCAGGACTCCAAACAGGCGGAAAAATGCTGCTACTCAGTTTTCTCTCTCCTCGGTGTGTT	480
Qy	481	GGGCATCTTTGGCCCTCACCTTCGCCCTTCATCATCGGACTGGACGGGAGCACAGGGCCCAAC	540
Db	481	GGGCATCTTTGGCCCTCACCTTCGCCCTTCATCATCGGACTGGACGGGAGCACAGGGCCCAAC	540
Qy	541	ACGCTTCTTCTCTTTGGGATCCCTCTTTTCCATCTGCTTCTCCTGCCTGCTGGCTCATGC	600
Db	541	ACGCTTCTTCTCTTTGGGATCCCTCTTTTCCATCTGCTTCTCCTGCCTGCTGGCTCATGC	600
Qy	601	TGTCAGTCTGACCAAGCTCGTCCGGGGGAGGAGCCCTTTCCCTGTGTGGTGAITCTGGG	660



```
QY 541 ACGTTCCTCTCTTTGGGATCCTCTTTTCAATCTGCTTCTCTGCTGCTGCTCATGC 600
Db 541 ACGTTCCTCTCTTTGGGATCCTCTTTTCAATCTGCTTCTCTGCTGCTGCTCATGC 600
QY 601 TGTCAGTCTGACCAAGCTCTGTCGGGGAGGAAGCCCTTTCCCTGTTGTTGATTTCTGGG 660
Db 601 TGTCAGTCTGACCAAGCTCTGTCGGGGAGGAAGCCCTTTCCCTGTTGTTGATTTCTGGG 660
QY 661 TCTGGCCGTGGGCTTCAGCTAGTCCAGGATGTTATCGCTATTGAATATATTGCTCTGAC 720
Db 661 TCTGGCCGTGGGCTTCAGCCCTAGTCCAGGATGTTATCGCTATTGAATATATTGCTCTGAC 720
QY 721 CATGAATAGGACCAACGTCATGTCCTTTCTGAGCTTTCCGCTCCCTCGTCGCAATGAAGA 780
Db 721 CATGAATAGGACCAACGTCATGTCCTTTCTGAGCTTTCCGCTCCCTCGTCGCAATGAAGA 780
QY 781 CTTTGTCTCTCTGCTCACCTACCTCTCTCTTTGATGGGCTGACCTTCTCATGTCCTC 840
Db 781 CTTTGTCTCTCTGCTCACCTACCTCTCTCTTTGATGGGCTGACCTTCTCATGTCCTC 840
QY 841 CTTACCTTCTGTTGCTTCTTACGGGCTGGAAGACATAGGGGCCCAACATCTACCTCAC 900
Db 841 CTTACCTTCTGTTGCTTCTTACGGGCTGGAAGACATAGGGGCCCAACATCTACCTCAC 900
QY 901 GATGCTCTCTCCATTTGCCATCTGGGTGGCTGGATCACCCCTGCTCATGCTTCTGACTT 960
Db 901 GATGCTCTCTCCATTTGCCATCTGGGTGGCTGGATCACCCCTGCTCATGCTTCTGACTT 960
QY 961 TGACCGAGGTGGATGACACCATCTCAGCTCCGCTTTGGCTTGGCCAAATGGCTGGGTGTT 1020
Db 961 TGACCGAGGTGGATGACACCATCTCAGCTCCGCTTTGGCTTGGCCAAATGGCTGGGTGTT 1020
QY 1021 CCGTTGGCTTATGTTAGTCCCGAGTTTGGCTGCTCACAAGAACGAAACCCCATGGA 1080
Db 1021 CCGTTGGCTTATGTTAGTCCCGAGTTTGGCTGCTCACAAGAACGAAACCCCATGGA 1080
QY 1081 TTATCCTGTTGAGGATGCTTTCTGTAACCTCAACTCGTGAAGAGAGCTATGGTGTGA 1140
Db 1081 TTATCCTGTTGAGGATGCTTTCTGTAACCTCAACTCGTGAAGAGAGCTATGGTGTGA 1140
QY 1141 GAAACAGAGCTACTCTCAAGAGGAAATCACTCAAGTTTGAAGAGACAGGGGACACGCT 1200
Db 1141 GAAACAGAGCTACTCTCAAGAGGAAATCACTCAAGTTTGAAGAGACAGGGGACACGCT 1200
QY 1201 CTATGCCCCCTATTCCACACATTTTTCAGCTGAGAACCCGCTCCCAAGAAAGAAATTC 1260
Db 1201 CTATGCCCCCTATTCCACACATTTTTCAGCTGAGAACCCGCTCCCAAGAAAGAAATTC 1260
QY 1261 CATCCACAGGGCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAGAAAGAGG 1320
Db 1261 CATCCACAGGGCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAGAAAGAGG 1320
QY 1321 CAGCTAACTCTGTCCTGAAGAGTGGGACAAATGACGCGGGCGGAGATCTAGCGGGAGC 1380
Db 1321 CAGCTAACTCTGTCCTGAAGAGTGGGACAAATGACGCGGGCGGAGATCTAGCGGGAGC 1380
QY 1381 TCAAGGGATGTCGGGGAATCTTGTAGTCTTCTGAGAAAACCTGTACAGACACTACGGGA 1440
Db 1381 TCAAGGGATGTCGGGGAATCTTGTAGTCTTCTGAGAAAACCTGTACAGACACTACGGGA 1440
QY 1441 ACAGTTTGCTCCTCCAGCTCAACCAAAATTTCCATGCTGGGGCTGATGTTGGCT 1500
Db 1441 ACAGTTTGCTCCTCCAGCTCAACCAAAATTTCCATGCTGGGGCTGATGTTGGCT 1500
QY 1501 AGTAAGACTCCAGTTCTTAGAGCGCTGTAGTATTTTTTTTTTTTTTTTGTCTCATCTTTGG 1560
Db 1501 AGTAAGACTCCAGTTCTTAGAGCGCTGTAGTATTTTTTTTTTTTTTTTGTCTCATCTTTGG 1560
QY 1561 ATACTTCTTTTAAAGTGGAGTCTCAGGCAACTCAAGTTTAGACCCCTACTCTTTTGTGTT 1620
Db 1561 ATACTTCTTTTAAAGTGGAGTCTCAGGCAACTCAAGTTTAGACCCCTACTCTTTTGTGTT 1620
QY 1621 GTTTTTTGAACAGGATCTTGTCTCTGTACCCAGGCTTGAGTGCAGTGGTGCGATCACAG 1680
```

```
Db 1621 GTTTTTTGAACAGGATCTTGTCTCTGTACCCAGGCTTGAGTGCAGTGGTGCGATCACAG 1680
QY 1681 CCAGTGCAGCTTCGACACCTGTGTCTCAAGCAATCTCTCCATCTCCATCTCCCAAAGTG 1740
Db 1681 CCAGTGCAGCTTCGACACCTGTGTCTCAAGCAATCTCTCCATCTCCATCTCCCAAAGTG 1740
QY 1741 CTGGGATGACAGGCGTAGGCCACAGCTCCAGCTAGGCCCTTAATCTTCTGCTTATTTT 1800
Db 1741 CTGGGATGACAGGCGTAGGCCACAGCTCCAGCTAGGCCCTTAATCTTCTGCTTATTTT 1800
QY 1801 CCATGACTTAAAGGCTTGTGTCTCAGCTCAGCTGGGTCAACAGCTCTAGGGGCGCTG 1860
Db 1801 CCATGACTTAAAGGCTTGTGTCTCAGCTCAGCTGGGTCAACAGCTCTAGGGGCGCTG 1860
QY 1861 CTCCTCTAACTACAGTGGGTTTTGTGAGGCTCTGTGGCCAGAGCAGACCTGCATATCT 1920
Db 1861 CTCCTCTAACTACAGTGGGTTTTGTGAGGCTCTGTGGCCAGAGCAGACCTGCATATCT 1920
QY 1921 GAGCAAAATAGCAAAAGCTCTCTCAGGCCACTGGCCCTGAATCTACACTGGAAGCCAAC 1980
Db 1921 GAGCAAAATAGCAAAAGCTCTCTCAGGCCACTGGCCCTGAATCTACACTGGAAGCCAAC 1980
QY 1981 TTGCTGGCACCCCGCTCCCAACCCCTTCTTGCTGGTAGGAGAGGCTAAAGATCACCC 2040
Db 1981 TTGCTGGCACCCCGCTCCCAACCCCTTCTTGCTGGTAGGAGAGGCTAAAGATCACCC 2040
QY 2041 TAAATTTACTCATCTCTTAGTGTCCCTCAATTTGGGCTCAGAGCTCCCCAGACCA 2100
Db 2041 TAAATTTACTCATCTCTTAGTGTCCCTCAATTTGGGCTCAGAGCTCCCCAGACCA 2100
QY 2101 ATTACAGGTCACCCCTCTCTTCTGCACTGCCCCAACTTGTCTCAATTCGAGATC 2160
Db 2101 ATTACAGGTCACCCCTCTCTTCTGCACTGCCCCAACTTGTCTCAATTCGAGATC 2160
QY 2161 TAATCTCCCTACGCTCTGCCAGGAATTTCTTTTCAGACCTCACTAGCACAAAGCCCGTTG 2220
Db 2161 TAATCTCCCTACGCTCTGCCAGGAATTTCTTTTCAGACCTCACTAGCACAAAGCCCGTTG 2220
QY 2221 CTCCTTGTGAGGAAATTTGTAGATCAATTTCTCAATTTCTGGGCTGATCTTCT 2280
Db 2221 CTCCTTGTGAGGAAATTTGTAGATCAATTTCTCAATTTCTGGGCTGATCTTCT 2280
QY 2281 CTCATCTTCACCCCAACCTCTGTAATAGATTTACCGATTTACGCTGCAATTCGTAA 2340
Db 2281 CTCATCTTCACCCCAACCTCTGTAATAGATTTACCGATTTACGCTGCAATTCGTAA 2340
QY 2341 GTGGGATGCTCTCTTAATGGAGAGTGTTCATTTGTATTAATAGTTATTACCTGAGTAT 2400
Db 2341 GTGGGATGCTCTCTTAATGGAGAGTGTTCATTTGTATTAATAGTTATTACCTGAGTAT 2400
QY 2401 GCAATAAAGATGTGTGGCCTCTTTTCATGTTGGTGGCAGCAAAAAA 2456
Db 2401 GCAATAAAGATGTGTGGCCTCTTTTCATGTTGGTGGCAGCAAAAAA 2456
```

## RESULT 7

AEA00088  
ID AEA00088 standard; cDNA; 2456 BP.

XX AEA00088;

XX 28-JUL-2005 (first entry)

XX Human TAT115 cDNA SEQ ID NO:40.

DE  
XX  
XX  
KW gene; tumor-associated antigen; cytostatic; breast tumor;  
endocrine disease; gynecology and obstetrics; neoplasm; colon tumor;  
gastrointestinal disease; rectal tumor; endometroid carcinoma;  
KW genitourinary disease; renal tumor; lung tumor; respiratory disease;  
ovary tumor; skin tumor; liver tumor.  
XX  
OS Homo sapiens.







Db 1201 CTATGCCCTTATCCACATATTTTCAGCTGCAGAACACGCTCCCCAAAAAGGAATTC 1260  
Qy 1261 CATCCACGGCCCGCTTGGCCGAGCCTTACAAGACTATGAAGTAAAGAAAGGG 1320  
Db 1261 CATCCACGGCCCGCTTGGCCGAGCCTTACAAGACTATGAAGTAAAGAAAGGG 1320  
Qy 1321 CAGCTAACTCTGTCTGAAGAGTGGGACAAATGCAGCCGGCGGCGAGATCTAGCGGAGC 1380  
Db 1321 CAGCTAACTCTGTCTGAAGAGTGGGACAAATGCAGCCGGCGGCGAGATCTAGCGGAGC 1380  
Qy 1381 TCAAGGGATGTGGGGAATCTTGAGTCTTCTGAGAAAACCTGTAAGAAGTAAAGAAAGGG 1440  
Db 1381 TCAAGGGATGTGGGGAATCTTGAGTCTTCTGAGAAAACCTGTAAGAAGTAAAGAAAGGG 1440  
Qy 1441 ACAGTTTGCCTCCCTCCAGCCTCAACCAAAATCTTCCATGCTGGGCTGATGGGCT 1500  
Db 1441 ACAGTTTGCCTCCCTCCAGCCTCAACCAAAATCTTCCATGCTGGGCTGATGGGCT 1500  
Qy 1501 AGTAAGACTCCAGTCTTCTAGAGCGCTGTAGTATTTTTTTTTTTTGTCTCAFCCTTTGG 1560  
Db 1501 AGTAAGACTCCAGTCTTCTAGAGCGCTGTAGTATTTTTTTTTTTTGTCTCAFCCTTTGG 1560  
Qy 1561 ATACTTTCTTTTAAAGTGGAGTCTCAGGCAACTCAAGTTTAGACCTTACTCTTTTGT 1620  
Db 1561 ATACTTTCTTTTAAAGTGGAGTCTCAGGCAACTCAAGTTTAGACCTTACTCTTTTGT 1620  
Qy 1621 GTTTTTTGAACAGGATCTTGCTCTCTCACCAGGCTTGAGTCAGTGGGATCACAG 1680  
Db 1621 GTTTTTTGAACAGGATCTTGCTCTCTCACCAGGCTTGAGTCAGTGGGATCACAG 1680  
Qy 1681 CCAGTGCAGCCTCGACACCTCTGTCTCAAGCAATCTCCCATCTCCATCTCCCAAGTG 1740  
Db 1681 CCAGTGCAGCCTCGACACCTCTGTCTCAAGCAATCTCCCATCTCCATCTCCCAAGTG 1740  
Qy 1741 CTGGGATGACAGCGTGAGCCACAGCTCCAGCCTTAGGCCCTTAAATCTTGTGTTATTT 1800  
Db 1741 CTGGGATGACAGCGTGAGCCACAGCTCCAGCCTTAGGCCCTTAAATCTTGTGTTATTT 1800  
Qy 1801 CCATGACTTAAAGTCTGCTCATCTCAGCTCAGCTGGCTCAGACGCTTAGGGGCTG 1860  
Db 1801 CCATGACTTAAAGTCTGCTCATCTCAGCTCAGCTGGCTCAGACGCTTAGGGGCTG 1860  
Qy 1861 CTCCTCTAACTCAGCTGGGTTTGTGAGGCTCTGTGGCCAGAGCAGACCTGCATATCT 1920  
Db 1861 CTCCTCTAACTCAGCTGGGTTTGTGAGGCTCTGTGGCCAGAGCAGACCTGCATATCT 1920  
Qy 1921 GAGCAAAAATAGCAAAAGCCTCTCTCAGCCCACTGGCCTGAATCTACACTGGAAGCCAAC 1980  
Db 1921 GAGCAAAAATAGCAAAAGCCTCTCTCAGCCCACTGGCCTGAATCTACACTGGAAGCCAAC 1980  
Qy 1981 TTGCTGGCACCCCGCTCCCAACCTTCTTGCTGGTAGGAGGCTAAAGATCACCC 2040  
Db 1981 TTGCTGGCACCCCGCTCCCAACCTTCTTGCTGGTAGGAGGCTAAAGATCACCC 2040  
Qy 2041 TAAATTTACTCATCTCTAGTCTGCTCACAATGGGCTCAGAGCTCCCGAGCACA 2100  
Db 2041 TAAATTTACTCATCTCTAGTCTGCTCACAATGGGCTCAGAGCTCCCGAGCACA 2100  
Qy 2101 ATTACAGGTCACCCCTCTCTTCTGACCTGTGCTCCCAAACTGCTGTCAATCCGAGATC 2160  
Db 2101 ATTACAGGTCACCCCTCTCTTCTGACCTGTGCTCCCAAACTGCTGTCAATCCGAGATC 2160  
Qy 2161 TAATCTCCCTAGCTCTGCGAGGAATCTTTTCAGACCTCAGTACAGCAAGCCGGTGG 2220  
Db 2161 TAATCTCCCTAGCTCTGCGAGGAATCTTTTCAGACCTCAGTACAGCAAGCCGGTGG 2220  
Qy 2221 CTCCTTGTGAGGAGATTTCTAGATCATCTCTCACTTCAAAATCTCTGGGCTGATCTCT 2280  
Db 2221 CTCCTTGTGAGGAGATTTCTAGATCATCTCTCACTTCAAAATCTCTGGGCTGATCTCT 2280  
Qy 2281 CTCATCTTGACCCCAACCTCTGTAAATAGATTTACCGCATTTACGGCTGCATTTGTAA 2340  
Db 2281 CTCATCTTGACCCCAACCTCTGTAAATAGATTTACCGCATTTACGGCTGCATTTGTAA 2340

Qy 2341 GTGGCATGCTCCTTAATGGAGGAGTTCATTGTATATAGTTATTCCACCTGAGTAT 2400  
Db 2341 GTGGCATGCTCCTTAATGGAGGAGTTCATTGTATATAGTTATTCCACCTGAGTAT 2400  
Qy 2401 GCAATAAAGATGTGGTGGCCACTCTTTTCATGTGGTGGCAGCAAAAAA 2456  
Db 2401 GCAATAAAGATGTGGTGGCCACTCTTTTCATGTGGTGGCAGCAAAAAA 2456

RESULT 8  
AEA00608  
ID AEA00608 standard; cDNA; 2456 BP.  
XX AEA00608;  
AC AEA00608;  
XX DT 28-JUL-2005 (first entry)  
XX Human TAT115 cDNA sequence SeqID40.  
XX antibody identification; tumor-associated antigen; cytostatic;  
KW antiference; gene therapy; cell death; cancer; breast tumor;  
KW colon tumor; rectal tumor; renal tumor; lung tumor; ovary tumor;  
KW skin tumor; liver tumor; gene; ss; TAT115.  
XX OS Homo sapiens.  
XX US2005107595-A1.  
XX PD 19-MAY-2005.  
XX PF 10-SEP-2004; 2004US-00938061.  
XX PR 20-JUN-2001; 2001US-0299500P.  
PR 29-JUN-2001; 2001US-0301880P.  
PR 18-SEP-2001; 2001US-0323268P.  
PR 19-JUN-2002; 2002US-00177488.  
PR 26-MAR-2004; 2004US-0557116P.  
PR 04-AUG-2004; 2004US-0598899P.  
XX (GETH ) GENENTECH INC.  
XX Cairns B, Chen R, Frantz G, Hillan KJ, Koepfen H, Phillips HS;  
PI Polakis P, Spencer SD, Smith V, Williams PM, Wu TD, Zhang Z;  
PI Sakanaka C, Chuntharapai A, Reed CJ;  
XX WPI; 2005-371577/38.  
DR P-PSDB; AEA00686.  
XX Novel isolated antibody e.g. anti-E16 or anti-TAT112 antibody that binds  
PT to tumor-associated antigenic target polypeptide, useful for diagnosing  
PT or treating cancer.  
XX Claim 1; SEQ ID NO 40; 96pp; English.  
XX This invention relates to a novel isolated antibody, for example anti-E16  
or anti-TAT112 antibody, that binds to a tumor-associated antigenic  
CC target polypeptide (TAT) and that lacks an associated signal peptide  
CC sequence. The invention may be useful for the development of compounds  
CC with a cytostatic activity acting as antagonists of the TAT118  
CC polypeptide or RNA interference whilst the disclosed sequences may be  
CC useful for gene therapy. The invention is useful for inducing the death  
CC of a cell (such as a cancer cell chosen from breast, colon, rectum,  
CC endometrium, kidney, lung, ovary, skin and liver) to which it binds,  
CC inhibiting proliferation or promoting cell death of a cell expressing  
CC TAT118. In addition, the invention may be useful for detecting the level  
CC of TAT118 polypeptide in a test cell relative to a control cell, or  
CC treating or preventing a cell proliferative disorder associated with  
CC increased expression of TAT118. The novel antibody of the invention is  
CC useful for inhibiting the growth of a cancer cell and may be useful for  
CC diagnosing or treating cancer. The present sequence is that of the human  
CC TAT115 cDNA which encodes a protein against which an antibody of the  
CC invention may be targeted.

XX										
SQ	Sequence	2456 BP; 519 A; 722 C; 571 G; 644 T; 0 U; 0 Other;								
QY	Query Match	100.0%; Score 2456; DB 14; Length 2456;								
	Best Local Similarity	100.0%; Pred. No. 0;								
	Matches 2456; Conservative	0; Mismatches	0; Indels	0; Gaps	0;					
QY	1	ATAACAGCATGAAGTGCCTTGGAATAGCGTGTCTCTCTCCCTCGAACCTCCCCCC	60							
Db	1	ATAACAGCATGAAGTGCCTTGGAATAGCGTGTCTCTCTCCCTCGAACCTCCCCCC	60							
QY	61	TCCTTGTCCCTCTGCTCACCCCTCGTTCCTCTCCCTCGGAGAGGGCGCCTTTATA	120							
Db	61	TCCTTGTCCCTCTGCTCACCCCTCGTTCCTCTCCCTCGGAGAGGGCGCCTTTATA	120							
QY	121	ACAACTGCTCAGAGTGCAGGGGGGATAGCTGTCGAAGGTCTCCCCAGCACTGAGGAG	180							
Db	121	ACAACTGCTCAGAGTGCAGGGGGGATAGCTGTCGAAGGTCTCCCCAGCACTGAGGAG	180							
QY	181	CTCGCCTGCTGCCCTCTTTGGCGGGGAAGCAGCACCAAGTTACGGCCCAACGCTTTGGC	240							
Db	181	CTCGCCTGCTGCCCTCTTTGGCGGGGAAGCAGCACCAAGTTACGGCCCAACGCTTTGGC	240							
QY	241	ACTAGGTCAGAAATGGCTACACAGTCCCTGATGGTTGCGCGCAATGGCCTGAATCCAA	300							
Db	241	ACTAGGTCAGAAATGGCTACACAGTCCCTGATGGTTGCGCGCAATGGCCTGAATCCAA	300							
QY	301	GTACTACAGACTTTGTGATAAGGCTGAAGCTTGGGGCATCGTCTGAACCGTGGCCAC	360							
Db	301	GTACTACAGACTTTGTGATAAGGCTGAAGCTTGGGGCATCGTCTGAACCGTGGCCAC	360							
QY	361	AGCGGGGTGTGACCTCGGTGGGCTTTCATGCTCACTCTCCGATCCTCGTCTGCAAGGT	420							
Db	361	AGCGGGGTGTGACCTCGGTGGGCTTTCATGCTCACTCTCCGATCCTCGTCTGCAAGGT	420							
QY	421	GCAGGACTCCAAACAGGCGAAATGCTGCTACTCAGTTTCTCTCTCTCGGTGTGTT	480							
Db	421	GCAGGACTCCAAACAGGCGAAATGCTGCTACTCAGTTTCTCTCTCTCGGTGTGTT	480							
QY	481	GGGCATCTTTGGCCTCACCTTCGCCTTCATCATCGGACTGGAGGGAGCACAGGGCCAC	540							
Db	481	GGGCATCTTTGGCCTCACCTTCGCCTTCATCATCGGACTGGAGGGAGCACAGGGCCAC	540							
QY	541	ACGCTTCTCTCTTTGGGATCCTCTTTTCCATCTGCTTCTCTCGCTGCTGGCTCATGC	600							
Db	541	ACGCTTCTCTCTTTGGGATCCTCTTTTCCATCTGCTTCTCTCGCTGCTGGCTCATGC	600							
QY	601	TGTCAGTCTGACCAAGCTCGTCCGGGGAGGAAGCCCTTTTCCCTGTTGGTGAATCTGGG	660							
Db	601	TGTCAGTCTGACCAAGCTCGTCCGGGGAGGAAGCCCTTTTCCCTGTTGGTGAATCTGGG	660							
QY	661	TCGCGCGTGGGCTTCAGCCTAGTCCAGGATGTATCGCTATTGAATATATTGTCTCGAC	720							
Db	661	TCGCGCGTGGGCTTCAGCCTAGTCCAGGATGTATCGCTATTGAATATATTGTCTCGAC	720							
QY	721	CATGAATAGGACCAACGTCAATGTCTTTTCTGAGCTTTTCGCTCCTCGTCGCAATGAAGA	780							
Db	721	CATGAATAGGACCAACGTCAATGTCTTTTCTGAGCTTTTCGCTCCTCGTCGCAATGAAGA	780							
QY	781	CTTTGTCTCTGCTCACCTACGTCCTCTTTCTTGATGGCGGTGACCTTCTCATGTCTCTC	840							
Db	781	CTTTGTCTCTGCTCACCTACGTCCTCTTTCTTGATGGCGGTGACCTTCTCATGTCTCTC	840							
QY	841	CTTTCACCTTCTGTGGTTTCCTTTCACGGGCTGGAAGAGACATGGGGCCACATCTACCTCAC	900							
Db	841	CTTTCACCTTCTGTGGTTTCCTTTCACGGGCTGGAAGAGACATGGGGCCACATCTACCTCAC	900							
QY	901	GATGCTCCTCTCCATTTGCCATCTGGGTGGCCTGGATCACCTGCTCATGCTTCTGACTTT	960							
Db	901	GATGCTCCTCTCCATTTGCCATCTGGGTGGCCTGGATCACCTGCTCATGCTTCTGACTTT	960							
QY	961	TGACCGCAGGTGGGATGACACCATCTCAGCTCCGCTTTGGCTTGGCAATGGCTGGGTGTT	1020							

961	TCACCGCAGGTGGGATGACACCATCTCAGCTCCGCTTGGCTGCCAATGGCTGGGTGTT	1020
1021	CCTGTTGGCTTATGTTAGTCCCGAGTTTGGCTGTCTCAAAAGCAACGAAACCCCATGGA	1080
1021	CCTGTTGGCTTATGTTAGTCCCGAGTTTGGCTGTCTCAAAAGCAACGAAACCCCATGGA	1080
1081	TTATCCTGTTGAGGATGCTTTCTGTAAACCTCAACTCGTGTGAAGAGAGCTATGGTGTGA	1140
1081	TTATCCTGTTGAGGATGCTTTCTGTAAACCTCAACTCGTGTGAAGAGAGCTATGGTGTGA	1140
1141	GAAACAGAGCTACTCTCAAGAGGAAATCACTCAAGGTTTGAAGAGAGAGGACACGCT	1200
1141	GAAACAGAGCTACTCTCAAGAGGAAATCACTCAAGGTTTGAAGAGAGAGGACACGCT	1200
1201	CTATGCCCTTATTCACACATTTTCAGCTGCAGAACCAAGCCTCCCCAAAAGGAATTC	1260
1201	CTATGCCCTTATTCACACATTTTCAGCTGCAGAACCAAGCCTCCCCAAAAGGAATTC	1260
1261	CATCCACGGGCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAGAAAGGG	1320
1261	CATCCACGGGCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAGAAAGGG	1320
1321	CAGCTAACTCTGCTGAAGAGTGGGACAAATGCAGCGGGCGGAGATCTAGCGGAGC	1380
1321	CAGCTAACTCTGCTGAAGAGTGGGACAAATGCAGCGGGCGGAGATCTAGCGGAGC	1380
1381	TCAAAGGAGTATGGGCGAAATCTTGAGTCTCTGAGAAAACGTGACAGACACTACGGGA	1440
1381	TCAAAGGAGTATGGGCGAAATCTTGAGTCTCTGAGAAAACGTGACAGACACTACGGGA	1440
1441	ACAGTTTGCCTCCCTCCAGCCTCAACCAATCTTCCATGCTGGGGCTGATGGGCT	1500
1441	ACAGTTTGCCTCCCTCCAGCCTCAACCAATCTTCCATGCTGGGGCTGATGGGCT	1500
1501	AGTAAGACTCCAGTCTTTAGAGCGCTGTAGTATTTTTTTTTTTTGTCTCATCTCTTTGG	1560
1501	AGTAAGACTCCAGTCTTTAGAGCGCTGTAGTATTTTTTTTTTTTGTCTCATCTCTTTGG	1560
1561	ATACTCTTTTAAAGTGGGAGTCTCAGGCAACTCAAGTTTAGACCCCTTACTCTTTTGT	1620
1561	ATACTCTTTTAAAGTGGGAGTCTCAGGCAACTCAAGTTTAGACCCCTTACTCTTTTGT	1620
1621	GTTTTTTGAACAGAGACTTGTCTGTACCCAGGCTTGAGTGCAGTGGGATCACAG	1680
1621	GTTTTTTGAACAGAGACTTGTCTGTACCCAGGCTTGAGTGCAGTGGGATCACAG	1680
1681	CCAGTGCAGCCTCGACACCTGTCTCAAGCAATCCTCCCATCTCCATCTCCCAAAGTG	1740
1681	CCAGTGCAGCCTCGACACCTGTCTCAAGCAATCCTCCCATCTCCATCTCCCAAAGTG	1740
1741	CTGGGATGACAGGCGTGAGCCACAGCTCCAGCCTAGGCCCTTAAATCTTGTGTTATTT	1800
1741	CTGGGATGACAGGCGTGAGCCACAGCTCCAGCCTAGGCCCTTAAATCTTGTGTTATTT	1800
1801	CCATGAGCTAAAGGCTTGGTCACTGAGCTACGCTGCTGCACACAGCTCTAGGGCCTG	1860
1801	CCATGAGCTAAAGGCTTGGTCACTGAGCTACGCTGCTGCACACAGCTCTAGGGCCTG	1860
1861	CTCCTCTAACTCACAAGTGGGTTTGTGAGGCTCTGTGGCCACAGAGACACTGTCATATCT	1920
1861	CTCCTCTAACTCACAAGTGGGTTTGTGAGGCTCTGTGGCCACAGAGACACTGTCATATCT	1920
1921	GAGCAAAAATAGCAAAAAGCCTCTCAGCCCACTGGCCTGAATCTACACTGGAAGCAAC	1980
1921	GAGCAAAAATAGCAAAAAGCCTCTCAGCCCACTGGCCTGAATCTACACTGGAAGCAAC	1980
1981	TTGCTGGCACCCCGCTCCCAACCTTCTTCCTGGGTAGGAGGCTTAAAGATCACCC	2040
1981	TTGCTGGCACCCCGCTCCCAACCTTCTTCCTGGGTAGGAGGCTTAAAGATCACCC	2040
2041	TAAATTTACTCATCTCTTAGTGTGCTTCAATTTGGGCTCAGCAGCTCCCGCAGACCA	2100
2041	TAAATTTACTCATCTCTTAGTGTGCTTCAATTTGGGCTCAGCAGCTCCCGCAGACCA	2100

```
QY 2101 ATTCACAGTCCACCCCTCTCTTGTGACATGTCCTCCCAAACTTGCTGTCAATTCGAGATC 2160
Db |||||||
QY 2101 ATTCACAGGTACCCCTCTCTTGTGACATGTCCTCCCAAACTTGCTGTCAATTCGAGATC 2160
Db |||||||
QY 2161 TAATCTCCCTTACGCTCTGCGAGGAATTTCTTTCAGACCTCACTAGCACAAAGCCCGTTG 2220
Db |||||||
QY 2161 TAATCTCCCTTACGCTCTGCGAGGAATTTCTTTCAGACCTCACTAGCACAAAGCCCGTTG 2220
Db |||||||
QY 2221 CTCTTGTGAGGAGAAATTTGTAGATCATCTTCACTTCAAAATTCCTGGGGCTGATCTTCT 2280
Db |||||||
QY 2221 CTCTTGTGAGGAGAAATTTGTAGATCATCTTCACTTCAAAATTCCTGGGGCTGATCTTCT 2280
Db |||||||
QY 2281 CTCACTTGTGACCCCAACCTCTGTAATAGATTTTACCGCATTTACGGCTGCAATTCGTAA 2340
Db |||||||
QY 2281 CTCACTTGTGACCCCAACCTCTGTAATAGATTTTACCGCATTTACGGCTGCAATTCGTAA 2340
Db |||||||
QY 2341 GTGGGATGCTCTCTTAATGAGGAGTGTTTCATTGTATATAAGTTATTCACCTGAGTAT 2400
Db |||||||
QY 2341 GTGGGATGCTCTCTTAATGAGGAGTGTTTCATTGTATATAAGTTATTCACCTGAGTAT 2400
Db |||||||
QY 2401 GCAATAAGATGTGGTGGCCACTCTTTCATGTTGGTGGCAGCAAAAAA 2456
Db |||||||
QY 2401 GCAATAAGATGTGGTGGCCACTCTTTCATGTTGGTGGCAGCAAAAAA 2456
Db |||||||

RESULT 9
AED47480
ID AED47480 standard; DNA; 2456 BP.
XX
AC AED47480;
XX
DT 15-DEC-2005 (first entry)
DE
DE Retinoic acid-inducible G-protein coupled receptor 3 gene, SEQ ID 2.
KW RNA interference; gene silencing;
KW retinoic acid-inducible G-protein coupled receptor 3; breast tumor;
KW cytostatic; cell growth; gene; ds.
XX
XX Homo sapiens.
OS
FH Key
FH Location/Qualifiers
FH CDS
FH 254..1327
FH /*tag= a
FH /product= "Retinoic acid-inducible G-protein coupled
FH receptor 3"
FH /*tag= b
FH /*tag= c
FH /*tag= d
FH /*tag= e
FH /*tag= f
FH /*tag= g
FH /*tag= h
FH /*tag= i
FH /*tag= j
FH misc_feature
FH 338..403
FH /*tag= b
FH /*tag= c
FH /*tag= d
FH /*tag= e
FH /*tag= f
FH /*tag= g
FH /*tag= h
FH /*tag= i
FH /*tag= j
FH misc_feature
FH 452..514
FH /*tag= c
FH /*tag= d
FH /*tag= e
FH /*tag= f
FH /*tag= g
FH /*tag= h
FH /*tag= i
FH /*tag= j
FH misc_feature
FH 497..1057
FH /*tag= d
FH /*tag= e
FH /*tag= f
FH /*tag= g
FH /*tag= h
FH /*tag= i
FH /*tag= j
FH misc_feature
FH 548..601
FH /*tag= e
FH /*tag= f
FH /*tag= g
FH /*tag= h
FH /*tag= i
FH /*tag= j
FH misc_feature
FH 656..718
FH /*tag= f
FH /*tag= g
FH /*tag= h
FH /*tag= i
FH /*tag= j
FH misc_feature
FH 776..832
FH /*tag= g
FH /*tag= h
FH /*tag= i
FH /*tag= j
FH misc_feature
FH 887..946
FH /*tag= h
FH /*tag= i
FH /*tag= j
FH polyA_signal
FH 2403..2408
FH /*tag= i
FH /*tag= j
FH polyA_site
FH 2442..2456
FH /*tag= j
FH
```

```
PN JP2005287456-A.
XX
PD 20-OCT-2005.
XX
PF 02-APR-2004; 2004JP-00110232.
XX
XX 02-APR-2004; 2004JP-00110232.
PR
XX 02-APR-2004; 2004JP-00110232.
XX
PA (NIPK ) NIPPON KAYAKU KK.
PA (UYNI-) UNIV NIPPON IKA.
XX
PI Nagahata T, Emi M;
XX
XX WPI: 2005-737982/76.
DR P-PSDB; AED47479.
XX
XX Substante that suppresses expression of retinoic acid-inducible G-protein
PT coupled receptor-3 RA13 gene or nucleic acid, useful for suppressing cell
PT growth, and as pharmaceutical or therapeutic agent of breast cancer.
XX
XX Claim 1; SEQ ID NO 2; 22pp; Japanese.
PS
XX The invention relates to a novel substance which suppresses the
CC expression of a retinoic acid-inducible G-protein coupled receptor (RA1)3
CC gene. The RA13 gene suppressing agent is useful for suppressing cell
CC growth, and as a pharmaceutical or therapeutic agent of breast cancer.
CC and is also useful in identifying a cell growth inhibitory substance. The
CC RA13 gene suppressing agent enables the identification of a cell growth
CC inhibitory substance. This polynucleotide sequence represents the
CC retinoic acid-inducible G-protein coupled receptor 3 gene of the
CC invention.
XX
XX Sequence 2456 BP; 519 A; 722 C; 571 G; 644 T; 0 U; 0 Other;
SQ
Query Match 100.0%; Score 2456; DB 14; Length 2456;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATAAACAGCATGAAGTCCCGTGAACCTGGAATAGGCGGTGCTCTCCCTCGACCTCCCC 60
Db |||||||
QY 61 TCCTTGTCCCTCTGCTCACCCCTCGCTTCCCTCCCGGAGGGCGGCTTTTATA 120
Db |||||||
QY 61 TCCTTGTCCCTCTGCTCACCCCTCGCTTCCCTCCCGGAGGGCGGCTTTTATA 120
Db |||||||
QY 121 ACAACTGTCTCAGAGTGCAGGGCGGATAGCTGTCCAAGTCTCCCGACACTGAGGAG 180
Db |||||||
QY 181 CTGCGCTGTGCGCTCTTTCGCGCGGGAAGCAGCACAAGTTACGGCCACCGCTTGGC 240
Db |||||||
QY 241 ACTAGGTCAGAAATGGCTTACAACTGCTGATGTTGCGCAATGGCTGAAATCCAA 300
Db |||||||
QY 301 GTACTACAGACTTTGTGATAAGCTGAAAGCTTGGGGCATCGTCTAGAAAACCGTGGCCAC 360
Db |||||||
QY 361 AGCGGGGTGTGACCTCGGTGGCTTTCATGCTCACTCTCCGATCCTCGTCTGCAAGGT 420
Db |||||||
QY 421 GCAGGACTCCAAACAGCGGAAATGCTGCTTCTCACTTCTTCTCTCTCTGCTGTT 480
Db |||||||
QY 481 GGGCATCTTTGGCCTCACCTTCGCTTTCATTCGGAAGTGGAGCGGAGCAGAGGGCCAC 540
Db |||||||
QY 481 GGGCATCTTTGGCCTCACCTTCGCTTTCATTCGGAAGTGGAGCGGAGCAGAGGGCCAC 540
Db |||||||
```

QY 541 ACCTTCTTCTTGGGATCCTCTTTTCCATCTGCTTCTCTGCTCTGCTCATCG 600  
DB |||||  
QY 541 ACGTCTCTTCTTGGGATCCTCTTTTCCATCTGCTTCTCTGCTCTGCTCATGC 600  
DB |||||  
QY 601 TGTCACTCTGACCAAGCTCGTCCGGGGAGGAGCCCTTCCCTGTTGGTGAATCTGGG 660  
DB |||||  
QY 601 TGTCACTCTGACCAAGCTCGTCCGGGGAGGAGCCCTTCCCTGTTGGTGAATCTGGG 660  
DB |||||  
QY 661 TCTGGCCGTGGGCTTCAGCCTAGTCAGGATGTTATCGCTATTGAATATAATTGTCCTGAC 720  
DB |||||  
QY 661 TCTGGCCGTGGGCTTCAGCCTAGTCAGGATGTTATCGCTATTGAATATAATTGTCCTGAC 720  
DB |||||  
QY 721 CATGAATAGGACCAACGCTCAATGTCTTTTCTGAGCTTTCGGTCTCTCGTCGCAATGAAGA 780  
DB |||||  
QY 721 CATGAATAGGACCAACGCTCAATGTCTTTTCTGAGCTTTCGGTCTCTCGTCGCAATGAAGA 780  
DB |||||  
QY 781 CTTTGTCTCTGCTCACCTACCTAGTCCTCTTCTGATGGCGCTGACCTTCTCATGTCCTC 840  
DB |||||  
QY 781 CTTTGTCTCTGCTCACCTAGTCCTCTTCTTGTATGGCGCTGACCTTCTCATGTCCTC 840  
DB |||||  
QY 841 CTTTCACTTCTGTGTTCTCTTACGGGCTGGAAGAGACATGGGGCCACATCTACCTCAC 900  
DB |||||  
QY 841 CTTTCACTTCTGTGTTCTCTTACGGGCTGGAAGAGACATGGGGCCACATCTACCTCAC 900  
DB |||||  
QY 901 GATGCTCTCTCATGTCATCTGGGTGGCTGGATCACCTGCTCATGCTTCTCTGACTT 960  
DB |||||  
QY 901 GATGCTCTCTCATGTCATCTGGGTGGCTGGATCACCTGCTCATGCTTCTCTGACTT 960  
DB |||||  
QY 961 TGACCCGAGTGGGATGACACCATCTCAGCTCCGCTTGGCTGGCCAAATGGCTGGGTGTT 1020  
DB |||||  
QY 961 TGACCCGAGTGGGATGACACCATCTCAGCTCCGCTTGGCTGGCCAAATGGCTGGGTGTT 1020  
DB |||||  
QY 1021 CTTGTTGGCTTATGTTAGTCCGAGTTTGGGTGCTCACAAGAACACGAAACCCCATGGA 1080  
DB |||||  
QY 1021 CTTGTTGGCTTATGTTAGTCCGAGTTTGGGTGCTCACAAGAACACGAAACCCCATGGA 1080  
DB |||||  
QY 1081 TTATCTCTGTGAGGATGCTTTCTGTAACCTCAACTCGTGAAGAGACTATGGTGTGA 1140  
DB |||||  
QY 1081 TTATCTCTGTGAGGATGCTTTCTGTAACCTCAACTCGTGAAGAGACTATGGTGTGA 1140  
DB |||||  
QY 1141 GAACAGAGCTACTCTCAAGAGAAATCACTCAAGGTTTGAAGACAGGGGACACGCT 1200  
DB |||||  
QY 1141 GAACAGAGCTACTCTCAAGAGAAATCACTCAAGGTTTGAAGACAGGGGACACGCT 1200  
DB |||||  
QY 1201 CTATGCCCTCTATTCACACATTTTCAGCTCGAGAACCGCTCCCCAAAAGGAATTCCT 1260  
DB |||||  
QY 1201 CTATGCCCTCTATTCACACATTTTCAGCTCGAGAACCGCTCCCCAAAAGGAATTCCT 1260  
DB |||||  
QY 1261 CATCCACGGGCCACGCTTGGCCGAGCCCTTACAAGACTATGAAGTAAAGAAAGAGG 1320  
DB |||||  
QY 1261 CATCCACGGGCCACGCTTGGCCGAGCCCTTACAAGACTATGAAGTAAAGAAAGAGG 1320  
DB |||||  
QY 1321 CAGCTAACTCTGCTCAAGAGTGGGACAAATGSCAGCCGGGCGAGATCTAGCGGAGC 1380  
DB |||||  
QY 1321 CAGCTAACTCTGCTCAAGAGTGGGACAAATGSCAGCCGGGCGAGATCTAGCGGAGC 1380  
DB |||||  
QY 1381 TCAAAAGGATGTGGGCGAAATCTTGAGTCTTCTGAGAAAACGTACAAGACACTACGGGA 1440  
DB |||||  
QY 1381 TCAAAAGGATGTGGGCGAAATCTTGAGTCTTCTGAGAAAACGTACAAGACACTACGGGA 1440  
DB |||||  
QY 1441 ACAGTTTGGCTCCCTCCGAGCTCAACCAATTCATGCTGGGCTGATGGGCT 1500  
DB |||||  
QY 1441 ACAGTTTGGCTCCCTCCGAGCTCAACCAATTCATGCTGGGCTGATGGGCT 1500  
DB |||||  
QY 1501 AGTAAGACTCCAGTCTTGAAGCGCTGTAGTATTTTTTTTTTTTGTCTCATCCCTTGG 1560  
DB |||||  
QY 1501 AGTAAGACTCCAGTCTTGAAGCGCTGTAGTATTTTTTTTTTTTGTCTCATCCCTTGG 1560  
DB |||||  
QY 1561 ATACTTCTTTTAAAGTGGGAGTCTCAGGCAACTCAAGTTTAGACCCCTTACTCTTTTGT 1620  
DB |||||  
QY 1561 ATACTTCTTTTAAAGTGGGAGTCTCAGGCAACTCAAGTTTAGACCCCTTACTCTTTTGT 1620  
DB |||||

QY 1621 GTTTTTGAAACAGGATCTTGCTCTGTCAACCAGGCTTGAGTGCAGTGGTCGATCACAG 1680  
DB |||||  
QY 1621 GTTTTTGAAACAGGATCTTGCTCTGTCAACCAGGCTTGAGTGCAGTGGTCGATCACAG 1680  
DB |||||  
QY 1681 CCCAGTGCAGCCTCGACACACCTGTCTCAAGCAATCCTCCCATCTCCATCTCCCAAAGTG 1740  
DB |||||  
QY 1681 CCCAGTGCAGCCTCGACACACCTGTCTCAAGCAATCCTCCCATCTCCATCTCCCAAAGTG 1740  
DB |||||  
QY 1741 CTGGGATGACAGCGGTGAGCCACAGCTCCAGCCTAGGCCCTTAAATCTTGTCTGTTATTT 1800  
DB |||||  
QY 1741 CTGGGATGACAGCGGTGAGCCACAGCTCCAGCCTAGGCCCTTAAATCTTGTCTGTTATTT 1800  
DB |||||  
QY 1801 CCATGGACTAAAGTCTTGTCATCTGAGCTCAGCTGGCTCACAAGCTCTAGGGCCCTG 1860  
DB |||||  
QY 1801 CCATGGACTAAAGTCTTGTCATCTGAGCTCAGCTGGCTCACAAGCTCTAGGGCCCTG 1860  
DB |||||  
QY 1861 CTCCTCTAACTCACAGTGGGTTTGTGAGGCTCTGTGGCCAGAGCAGACCTGTCATATCT 1920  
DB |||||  
QY 1861 CTCCTCTAACTCACAGTGGGTTTGTGAGGCTCTGTGGCCAGAGCAGACCTGTCATATCT 1920  
DB |||||  
QY 1921 GAGCAAAAATAGCAAAAGCCTCTCTAGCCCACTGGCCTGAAATCTTACACTGGAAGCAAC 1980  
DB |||||  
QY 1921 GAGCAAAAATAGCAAAAGCCTCTCTAGCCCACTGGCCTGAAATCTTACACTGGAAGCAAC 1980  
DB |||||  
QY 1981 TTGCTGGCACCCCGCTCCCAACCTTCTTCCCTGGGTAGGAGGCTTAAAGATCACCC 2040  
DB |||||  
QY 1981 TTGCTGGCACCCCGCTCCCAACCTTCTTCCCTGGGTAGGAGGCTTAAAGATCACCC 2040  
DB |||||  
QY 2041 TAAATTTACTCATCTCTTAGTGTGCTCACAATGGGCTCAGCAGCTCCCCAGCACCA 2100  
DB |||||  
QY 2041 TAAATTTACTCATCTCTTAGTGTGCTCACAATGGGCTCAGCAGCTCCCCAGCACCA 2100  
DB |||||  
QY 2101 ATTACAGTCAACCCCTCTCTTTCGACTGTCCCAAACTTGGCTGTCAAATCCGAGATC 2160  
DB |||||  
QY 2101 ATTACAGTCAACCCCTCTCTTTCGACTGTCCCAAACTTGGCTGTCAAATCCGAGATC 2160  
DB |||||  
QY 2161 TAAATCTCCCTTACGCTCTGCCAGGAATCTTTTTCAGACCTCAGTACGCAAGCCCGTTG 2220  
DB |||||  
QY 2161 TAAATCTCCCTTACGCTCTGCCAGGAATCTTTTTCAGACCTCAGTACGCAAGCCCGTTG 2220  
DB |||||  
QY 2221 CTCCTTGTGAGAGAAATTTGTAGATCATTTCTCATTCAAATTCCTGGGCTGATCTTCT 2280  
DB |||||  
QY 2221 CTCCTTGTGAGAGAAATTTGTAGATCATTTCTCATTCAAATTCCTGGGCTGATCTTCT 2280  
DB |||||  
QY 2281 CTCATCTTCACACCCCAACCTCTGTAATAGATTTACCGATTTACGGCTGCATTCTGTAA 2340  
DB |||||  
QY 2281 CTCATCTTCACACCCCAACCTCTGTAATAGATTTACCGATTTACGGCTGCATTCTGTAA 2340  
DB |||||  
QY 2341 GTGGCATGCTCTCCTTAATGGAGGAGTGTTCATTGTATAATAGTTATTACCTGAGTAT 2400  
DB |||||  
QY 2341 GTGGCATGCTCTCCTTAATGGAGGAGTGTTCATTGTATAATAGTTATTACCTGAGTAT 2400  
DB |||||  
QY 2401 GCAATAAAGATGTGGTGGCCACTCTTTTCATGTTGGTGGCAGCAAAAAA 2456  
DB |||||  
QY 2401 GCAATAAAGATGTGGTGGCCACTCTTTTCATGTTGGTGGCAGCAAAAAA 2456  
DB |||||  
RESULT 10  
ADI28525  
ID ADI28525 standard; cdNA; 2456 BP.  
XX  
AC ADI28525;  
XX  
DT 22-APR-2004 (first entry)  
XX  
XX Human GPCR retinoic acid induced 3 (RAI-3) cdNA.  
DE  
XX Retinoic acid induced 3; RAI-3; human; G-protein coupled receptor; GPCR;  
KW antinflammatory; immunosuppressive; cytostatic; cardiant; aniallergic;  
KW broncholytic; gene therapy; gene; single nucleotide polymorphism; SNP;  
KW chromosome 12p13-p12.3; ss.  
XX  
OS Homo sapiens.

```

XX Key Location/Qualifiers
FH variation replace(112,r)
FT /*tag= a
FT /label= RAI-3-s1
FT /note= "located in 5' untranslated region"
FT /standard_name= "Single nucleotide polymorphism"
FT CDS 254..1327
FT /*tag= b
FT /product= "Human RAI3"
FT /replace(364,y)
FT /*tag= c
FT /label= RAI-3-s2
FT /note= "exon 1, silent (Ala/Ala)"
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(511,y)
FT /*tag= d
FT /label= RAI-3-s3
FT /note= "exon 2, silent (Ile/Ile)"
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(523,y)
FT /*tag= e
FT /label= RAI-3-s4
FT /note= "exon 2, silent (Asp/Asp)"
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(605,r)
FT /*tag= f
FT /label= RAI-3-s6
FT /note= "exon 2, missense (Ser/Gly)"
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(797,r)
FT /*tag= g
FT /label= RAI-3-s5
FT /note= "exon 2, missense (Thr/Ala)"
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(111,r)
FT /*tag= h
FT /label= RAI-3-s8
FT /note= "silent (Pro/Pro)"
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(1173,r)
FT /*tag= i
FT /label= RAI-3-s9
FT /note= "missense (Gln/Arg)"
FT /standard_name= "Single nucleotide polymorphism"
XX W02004001060-A2.
XX
XX 31-DEC-2003.
XX
XX 20-JUN-2003; 2003WO-US019255.
XX
XX 20-JUN-2002; 2002US-0390850P.
XX 29-AUG-2002; 2002US-0407006P.
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
XX Whitney GS, Opitck G, Garulacan L, Ramanathan CS, McKinnon M;
XX Bennett KL, Barber UE, Cacace A, Tsuchihashi Z;
XX WPI; 2004-090973/09.
XX P-PSDB; ADI28526.
XX
XX New nucleic acid molecule encoding a human G-protein coupled receptor,
XX RAI3, useful for preventing, treating or ameliorating chronic obstructive
XX pulmonary disease (COPD), COPD-like disorder, or the underlying symptoms
XX of COPD.
XX
XX Claim 21; SEQ ID NO 18; 301pp; English.
XX
XX The present sequence is that of cDNA encoding a human G-protein coupled
XX receptor (Class C, Group 5), denoted retinoic acid induced 3 (RAI-3).
XX Proteomics methods were used to isolate cigarette smoke-inducible

```

```

CC tyrosine phosphorylated proteins from airway epithelial cells. RAI-3 was
CC identified as being tyrosine phosphorylated and/or as being
CC associated/complexed with tyrosine phosphorylated proteins only in those
CC cells that had been exposed to cigarette smoke. Since RAI-3 is primarily
CC expressed in lung tissue, and since cigarette smoke is a major causative
CC factor of chronic obstructive pulmonary disease (COPD), RAI-3 provides a
CC novel cellular target for identifying modulators, e.g. agonists or
CC antagonists, useful for the treatment and/or prevention of COPD and
CC related disorders such as emphysema and chronic bronchitis. RAI-3
CC modulators, e.g. agonists and antagonists, especially antiseense
CC compounds, can be used to treat COPD and other disorders and diseases
CC associated with regulation of NF-kB and/or its associated or interacting
CC signaling molecules. Single nucleotide polymorphisms (SNPs) detected in
CC the RAI-3 gene are useful for determining COPD association in
CC individuals. RAI-3 nucleic acid molecules and polypeptides are useful for
CC preventing, treating or ameliorating disorders related to aberrant GPCR
CC signaling or cell cycle regulation, pulmonary disorders, inflammatory
CC lung disorders, COPD, the underlying symptoms of COPD, COPD-related
CC disorders or conditions, autoimmune disorders, disorders related to
CC hyperimmune activity, inflammatory conditions, disorders related to
CC aberrant acute phase responses, hypercongenital conditions, birth
CC defects, necrotic lesions, wounds, organ transplant rejection, renal
CC diseases, ischaemia-reperfusion injury, heart disorders, disorders
CC related to aberrant signal transduction, proliferation disorders,
CC cancers, HIV infection, asthma, cystic fibrosis, pulmonary fibrosis,
CC ulcerative colitis, cerebral infarct, myocardial infarct, diabetic
CC nephropathy, allergic rhinitis, Crohn's disease, atherosclerosis,
CC rheumatoid arthritis, inflammatory/autoimmune disorders, glioblastoma,
CC pulmonary small cell undifferentiated carcinoma, carcinoma of the breast,
CC colon, lung, ovary, pancreas, prostate, non-Hodgkin's lymphoma, disorders
CC associated with aberrant cell adhesion, I-CAM function and/or regulation,
CC E-selectin function and/or regulation, or aberrant NF-kB function and/or
CC regulation (all claimed).
XX
XX SQ Sequence 2456 BP; 516 A; 719 C; 570 G; 643 T; 0 U; 8 Other;
XX
XX Query Match 99.7%; Score 2448; DB 12; Length 2456;
XX Best Local Similarity 99.7%; Pred. No. 0;
XX Matches 2448; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
XX
XX QY 1 ATAACAGCATGAAGTCCGCTGGAACTGGAACTAGCGGTGCTCTCTCCCTCGACCCCTCCCC 60
XX DB 1 ATAACAGCATGAAGTCCGCTGGAACTGGAACTAGCGGTGCTCTCTCCCTCGACCCCTCCCC 60
XX
XX QY 61 TCCTTGTCCCTCTGCTCAGCCCTCGCTCGTTCCTCCCTCCGCGAGGGCGGCTTTATA 120
XX DB 61 TCCTTGTCCCTCTGCTCAGCCCTCGCTCGTTCCTCCCTCCGCGAGGGCGGCTTTATA 120
XX
XX QY 121 ACAACTGCTCAGAGTCGAGGGGGGATAGCTGTCCAAGGTCTCCCCAGCACTGAGGAG 180
XX DB 121 ACAACTGCTCAGAGTCGAGGGGGGATAGCTGTCCAAGGTCTCCCCAGCACTGAGGAG 180
XX
XX QY 181 CTCGCCCTGTGCCCTCTTTCGCGCGGGAAGCAGCACCAAGTTCACGGCCAAACGCTTGGC 240
XX DB 181 CTCGCCCTGTGCCCTCTTTCGCGCGGGAAGCAGCACCAAGTTCACGGCCAAACGCTTGGC 240
XX
XX QY 241 ACTAGGGTCCAGAAATGGCTACACAGTCCCTGATGTTGTCGCGCAATGGCTGAAATCCAA 300
XX DB 241 ACTAGGGTCCAGAAATGGCTACACAGTCCCTGATGTTGTCGCGCAATGGCTGAAATCCAA 300
XX
XX QY 301 GTACTACAGACTTTGTGATAAGCTCAAGCTTGGGGCATCGTCTAGAAACGGTGGCCAC 360
XX DB 301 GTACTACAGACTTTGTGATAAGCTTGGGGCATCGTCTAGAAACGGTGGCCAC 360
XX
XX QY 361 AGCGGGGTGTGACCTCGGTGGCCCTTCACTGCTCCTCTCCCGATCCTCGTCTGCAAGGT 420
XX DB 361 AGCGGGGTGTGACCTCGGTGGCCCTTCACTGCTCCTCTCCCGATCCTCGTCTGCAAGGT 420
XX
XX QY 421 GCGGACTCCACAGCGGAAAAATGCTGCCTACTCAGTTTCTCTTCTCTCTGCGGTGTT 480
XX DB 421 GCGGACTCCACAGCGGAAAAATGCTGCCTACTCAGTTTCTCTTCTCTCTGCGGTGTT 480
XX
XX QY 481 GGGCATCTTTGGGCTCACCTTCGCTTCATCGGACTGCGGAGGACGACAGGGCCAC 540

```



PN	US2003099974-A1.	Db	889	GGGATCTTTGGGCTCACCTTCGCTTTCATCATCGGACTGACGAGGACACAGGCCCCAC	948
XX					
PD	29-MAY-2003.	Qy	541	ACGCTTCTTCTCTTTGGGATCCTCTTTTCCATCTGCTCTCTCTGCTGCTGCTCATGC	600
XX					
PF	18-JUL-2002; 2002US-00198846.	Db	949	ACGCTTCTTCTCTTTGGGATCCTCTTTTCCATCTGCTCTCTCTGCTGCTGCTCATGC	1008
XX					
PR	18-JUL-2001; 2001US-0306220P.	Qy	601	TGTCAGTCTGACCAAGCTCGTCCGGGGAGGAGCCCTTTTCCCTGTTGGTGAATTCGGG	660
XX					
PA	(MILL-) MILLENNIUM PHARM INC.	Db	1009	TGTCAGTCTGACCAAGCTCGTCCGGGGAGGAGCCCTTTTCCCTGTTGGTGAATTCGGG	1068
XX					
PI	Lillie J, Xu Y, Wang Y, Steinmann K;	Qy	661	TCTGGCGGTGGGCTTCCAGCTAGTCAGGATGTTATCGCTATTGAAATATATTCCTGAC	720
XX					
DR	WPI; 2003-787014/74.	Db	1069	TCTGGCGGTGGGCTTCCAGCTAGTCAGGATGTTATCGCTATTGAAATATATTCCTGAC	1128
XX					
PT	Novel isolated polypeptide associated with breast cancer, useful for	Qy	721	CATGAATAGAACCAAGCTCAATCTCTTCTTGAGCTTTCGCTCCTCGTCGAATGAAGA	780
XX	detecting presence of polypeptide in sample, as a marker for breast	Db	1129	CATGAATAGAACCAAGCTCAATCTCTTCTTGAGCTTTCGCTCCTCGTCGAATGAAGA	1188
PT	cancer.				
XX					
XX	Disclosure; SEQ ID NO 10424; 36pp; English.	Qy	781	CTTTGTCTCTCTCTCACCTTACCTTCTCTTCTTGATGGCGCTGACCTTCTCTCATGTCTC	840
PS					
CC	The invention relates to an isolated polypeptide (I) associated with	Db	1189	CTTTGTCTCTCTCTCACCTTACCTTCTCTTCTTGATGGCGCTGACCTTCTCTCATGTCTC	1248
CC	breast cancer which is encoded by a nucleic acid molecule comprising a				
CC	nucleotide sequence (SI). Further disclosed is an antibody that binds to	Qy	841	CTTACCTTCTGTGGTCTCTTACGGGCTGGAAGAGACATGGGGCCACATCTACCTCAC	900
CC	the polypeptide of the invention. The activity of the polypeptide of the	Db	1249	CTTACCTTCTGTGGTCTCTTACGGGCTGGAAGAGACATGGGGCCACATCTACCTCAC	1308
CC	invention may be described as cytostatic. The antibody is useful for				
CC	detecting the presence of (I) in a sample. Nucleic acid molecules of the	Qy	901	GATGCTCTCTCTCATTTGGCATCTGGGTGGGCTGGATCACCCTCTCATGCTTCTGACTT	960
CC	invention are useful in the detection of breast tumours. (I) is useful as	Db	1309	GATGCTCTCTCTCATTTGGCATCTGGGTGGGCTGGATCACCCTCTCATGCTTCTGACTT	1368
CC	a marker for breast cancer and in breast cancer therapy. Sequences given	Qy	961	TGACCGCAGGTGGGATGACACCATCTCAGCTCCGCTTGGCTGCCAATGGCTGGGTGTT	1020
CC	in records ACN78851-ACN92934 represent nucleic acid markers associated	Db	1369	TGACCGCAGGTGGGATGACACCATCTCAGCTCCGCTTGGCTGCCAATGGCTGGGTGTT	1428
CC	with breast cancer. Note: The sequence listing does not form part of the	Qy	1021	CTTGTGGCTTATGTTAGTCCCGAGTTTGGCTGCACAAAGCAACCAACCCCATGGA	1080
CC	specification but may be obtained in electronic format from the USPTO web	Db	1429	CTTGTGGCTTATGTTAGTCCCGAGTTTGGCTGCACAAAGCAACCAACCCCATGGA	1488
CC	site at <a href="http://seqdata.uspto.gov/sequence.html?DocID=20030099974">seqdata.uspto.gov/sequence.html?DocID=20030099974</a>				
XX					
SQ	Sequence 4239 BP; 859 A; 1161 C; 1075 G; 1141 T; 0 U; 3 Other;	Qy	1081	TTATCCGTGTGAGGATGCTTCTGTAACCTCAACTCGTGAAGAGAGCTATGGTGGGA	1140
	Query Match 99.5%; Score 2443; DB 11; Length 4239;	Db	1489	TTATCCGTGTGAGGATGCTTCTGTAACCTCAACTCGTGAAGAGAGCTATGGTGGGA	1548
	Best Local Similarity 100.0%; Pred. No. 0;	Qy	1141	GAACAGAGCTACTCTCAAGAGGAAATCACTCAAGGTTTTTGAAGAGACAGGGACACGCT	1200
	Matches 2443; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Db	1549	GAACAGAGCTACTCTCAAGAGGAAATCACTCAAGGTTTTTGAAGAGACAGGGACACGCT	1608
Qy	1 ATACACGATGAAGTCGCTGGAACTGGAATAGGCTGCTCTCTCCCTGACCTCCCC	Qy	1201	CTATGCCCTTATTCACACATTTTTCAGCTGCAGAACCCAGCTTCCCCAAAGGAATTC	1260
Db	409 ATAACAGCATGAATGCGGTGGAACTGGAATAGGCTGCTCTCTCCCTGACCTCCCC	Db	1609	CTATGCCCTTATTCACACATTTTTCAGCTGCAGAACCCAGCTTCCCCAAAGGAATTC	1668
Qy	61 TCCTTGTCTCTGCTCACCCCTCGCTGTTTCTCTCCCTCCGCGAGGGCGCTTTATA	Qy	1261	CATCCCAAGGGCCACGCTTGGCGGAGCCCTTACAAAGACTATGAAAGTAAGAAAGGG	1320
Db	469 TCCTTGTCTCTGCTCACCCCTCGCTGTTTCTCTCCCTCCGCGAGGGCGCTTTATA	Db	1669	CATCCCAAGGGCCACGCTTGGCGGAGCCCTTACAAAGACTATGAAAGTAAGAAAGGG	1728
Qy	121 ACAACTGCTCAGAGTGGAGGGGGATAGCTGTCCAAGGTCTCCCCCAGCACTGAGGAG	Qy	1321	CAGCTAACTCTGTCTGAAGAGTGGGACAAATCAGCCGGCGGCGAGATCTAGCGGAGC	1380
Db	529 ACAACTGCTCAGAGTGGAGGGGGATAGCTGTCCAAGGTCTCCCCCAGCACTGAGGAG	Db	1729	CAGCTAACTCTGTCTGAAGAGTGGGACAAATCAGCCGGCGGCGAGATCTAGCGGAGC	1788
Qy	181 CTGCTGCTGCTCTTGGCGGGGGAGGAGCAAGTTCAGGGCCAAAGCTTGGC	Qy	1381	TCAAGGGATGTGGGCGAAATCTTGAGTCTTCTTGAGAAAACCTGTACAAGACACTAC	1440
Db	589 CTGCTGCTGCTCTTGGCGGGGGAGGAGCAAGTTCAGGGCCAAAGCTTGGC	Db	1789	TCAAGGGATGTGGGCGAAATCTTGAGTCTTCTTGAGAAAACCTGTACAAGACACTAC	1848
Qy	241 ACTAGGTCAGAAATGCTACAAAGCTCCTGTGTTGCGCAATGGCTGAAATCCAA	Qy	1441	ACAGTTTGGCTTCCCTCCAGCTTCAACCAATTTCTTCCATGTGGGGCTGATGTGGCT	1500
Db	649 ACTAGGTCAGAAATGCTACAAAGCTCCTGTGTTGCGCAATGGCTGAAATCCAA	Db	1849	ACAGTTTGGCTTCCCTCCAGCTTCAACCAATTTCTTCCATGTGGGGCTGATGTGGCT	1908
Qy	301 GTACTACAGACTTTGTGATAGGCTGAAGCTTGGGGCATCGTCTTAGAAAACGGTGGCCAC	Qy	1501	AGTAAGACTCCAGTCTTAGAGGCGCTGATGATTTTTTTTTTTTTTTTGTCTCATCTTGG	1560
Db	709 GTACTACAGACTTTGTGATAGGCTGAAGCTTGGGGCATCGTCTTAGAAAACGGTGGCCAC	Db	1909	AGTAAGACTCCAGTCTTAGAGGCGCTGATGATTTTTTTTTTTTTTTTGTCTCATCTTGG	1968
Qy	421 GCAGGACTCCAAAGCGGAAAATGCTGCTACTCAGTTTCTCTCTCTCTGGGTGTT	Qy	1561	ATACTTCTTTTAAAGTCGGAGTCTCAGGCAACTCAAGTTTAGACCCCTTACTCTTTTGT	1620
Db	829 GCAGGACTCCAAAGCGGAAAATGCTGCTACTCAGTTTCTCTCTCTCTGGGTGTT	Db	1969	ATACTTCTTTTAAAGTCGGAGTCTCAGGCAACTCAAGTTTAGACCCCTTACTCTTTTGT	2028
Qy	481 GGGCATCTTTGGCTCACCTTCGCTTTCATCATCGGACTGGAGGGAGCAGGGGCCAC				







301 GTACTCAGACTTTGTGATAAGGCTGAAGCTTGGGCGATCGTCTCTAGAAAACGCTGGCCAC 360  
301 GTACTCAGACTTTGTGATAAGGCTGAAGCTTGGGCGATCGTCTCTAGAAAACGCTGGCCAC 360  
361 AGCGGGGTTGTGACCTCGGTGGCTTTCATGCTCACTCTCCCGATCCTCGTCTGCAAGGT 420  
361 AGCGGGGTTGTGACCTCGGTGGCTTTCATGCTCACTCTCCCGATCCTCGTCTGCAAGGT 420  
421 GCAGGACTCAACAGGCGAAAATGCTCCCTACTCACTAGTTTCTCTCTCTCTGGTGTGTT 480  
421 GCAGGACTCAACAGGCGAAAATGCTCCCTACTCACTAGTTTCTCTCTCTCTGGTGTGTT 480  
481 GGGCATCTTTGGCTCACTTCGCTTTCATCATCGGACTGGAACGGGAGCACAGGGCCAC 540  
481 GGGCATCTTTGGCTCACTTCGCTTTCATCATCGGACTGGAACGGGAGCACAGGGCCAC 540  
541 ACGETTCTTCTCTTTGGGATCTCTTTTCCATCTGCTTCTCTCTGCTGCTGCTCATGC 600  
541 ACGETTCTTCTCTTTGGGATCTCTTTTCCATCTGCTTCTCTCTGCTGCTGCTCATGC 600  
601 TGTCACTCAACCAAGCTCTCGGGGAGGAGCCCTTTCCTGTTGGTGTATCTGGG 660  
601 TGTCACTCAACCAAGCTCTCGGGGAGGAGCCCTTTCCTGTTGGTGTATCTGGG 660  
661 TCTGGCGCTGGGCTTCAAGCTTAGTCCAGGATGTTATCGCTATTGAATATATTGCTCTGAC 720  
661 TCTGGCGCTGGGCTTCAAGCTTAGTCCAGGATGTTATCGCTATTGAATATATTGCTCTGAC 720  
721 CATGAATPAGGACCAACGTCATATGCTTTTCTGAGCTTTCGCTCTCTCGTGCATGAAGA 780  
721 CATGAATPAGGACCAACGTCATATGCTTTTCTGAGCTTTCGCTCTCTCGTGCATGAAGA 780  
781 CTTTGTCTCTCTGCTCACTACGCTCTCTCTTGTAGTGGCGCTGACCTTCTCATGCTCTC 840  
781 CTTTGTCTCTCTGCTCACTACGCTCTCTCTTGTAGTGGCGCTGACCTTCTCATGCTCTC 840  
841 CTTTCACTCTCTGTTGCTTCTTCAAGGCTGGAAGAGACATGGGGCCACATCTACCTCAC 900  
841 CTTTCACTCTCTGTTGCTTCTTCAAGGCTGGAAGAGACATGGGGCCACATCTACCTCAC 900  
901 GATGCTCTCTCATTTGCCATCTGGGTGGCTGGATCACCTCTCATGCTTCTGACTT 960  
901 GATGCTCTCTCATTTGCCATCTGGGTGGCTGGATCACCTCTCATGCTTCTGACTT 960  
961 TGACCGGAGTGGGATGACACCACTCTCAGCTCGGCTTGGCTGCCCAATGGCTGGGTGTT 1020  
961 TGACCGGAGTGGGATGACACCACTCTCAGCTCGGCTTGGCTGCCCAATGGCTGGGTGTT 1020  
1021 CTTGTTGGCTTATGTTAGTCCCGAGTTTGGCTGCTCACAAGCAACGAAACCCCATGGA 1080  
1021 CTTGTTGGCTTATGTTAGTCCCGAGTTTGGCTGCTCACAAGCAACGAAACCCCATGGA 1080  
1081 TTATCTGTTGAGGATGCTTCTGTAAACCTCAACTCGTGAAGAAGAGCTATGGTGGGA 1140  
1081 TTATCTGTTGAGGATGCTTCTGTAAACCTCAACTCGTGAAGAAGAGCTATGGTGGGA 1140  
1141 GAACAGAGCCTACTCTCAAGAGGAATCACTCAAGTTTGAAGAGACAGGAGACAGCT 1200  
1141 GAACAGAGCCTACTCTCAAGAGGAATCACTCAAGTTTGAAGAGACAGGAGACAGCT 1200  
1201 CTATGCCCCCTATTTCACACATTTTCACTGCTGAGAACCCCTCCCAAGGAATTTCTC 1260  
1201 CTATGCCCCCTATTTCACACATTTTCACTGCTGAGAACCCCTCCCAAGGAATTTCTC 1260  
1261 CATCCCAACGGGCCACGCTTGGCCGAGCCCTTACAAAGACTATCAAGTAAAGAAAGGG 1320  
1261 CATCCCAACGGGCCACGCTTGGCCGAGCCCTTACAAAGACTATCAAGTAAAGAAAGGG 1320  
1321 CAGCTAACTCTGCTCAAGAGTGGGACAAATGAGCCGGCGGCGAGATCTAGCGGAGC 1380  
1321 CAGCTAACTCTGCTCAAGAGTGGGACAAATGAGCCGGCGGCGAGATCTAGCGGAGC 1380  
1381 TCAAGGGATGTGGGCGAAATCTTGAGTCTTCTGAGAAAACCTGTACAAGACACTACGGGA 1440

1381 TCAAGGGATGTGGGCGAAATCTTGAGTCTTCTGAGAAAACCTGTACAAGACACTACGGGA 1440  
1441 ACAGTTTGGCTCCCTCCAGCTCAACCAATTTCTTCATGCTGGGGCTGATGTGGCT 1500  
1441 ACAGTTTGGCTCCCTCCAGCTCAACCAATTTCTTCATGCTGGGGCTGATGTGGCT 1500  
1501 AGTAAGACTCCAGTCTTAGAGGGCTGTAGTATTTTTTTTTTTTTTTTGTCTCATCTTTGG 1560  
1501 AGTAAGACTCCAGTCTTAGAGGGCTGTAGTATTTTTTTTTTTTTTTTGTCTCATCTTTGG 1560  
1561 ATACTTCTTTTAAAGTGGGAGTCTCAGGCAACTCAAGTTTAGACCTTTACTCTTTTCTTT 1620  
1561 ATACTTCTTTTAAAGTGGGAGTCTCAGGCAACTCAAGTTTAGACCTTTACTCTTTTCTTT 1620  
1621 GTTTTTTGAACAAGGATCTTGTCTCACCCAGGCTTGAGTGCAGTGCATCACAG 1680  
1621 GTTTTTTGAACAAGGATCTTGTCTCACCCAGGCTTGAGTGCAGTGCATCACAG 1680  
1681 CCCAGTGCAGCTCGACCACTGTGTCTCAAGCAATCTCCCATCTCCCATCTCCCAAGTG 1740  
1681 CCCAGTGCAGCTCGACCACTGTGTCTCAAGCAATCTCCCATCTCCCATCTCCCAAGTG 1740  
1741 CTGGGATGACAGGCTGAGCCACAGCTCCCAGCTTAGGCCCTTAAATCTTCTGTATTTT 1800  
1741 CTGGGATGACAGGCTGAGCCACAGCTCCCAGCTTAGGCCCTTAAATCTTCTGTATTTT 1800  
1801 CCATGACTTAAAGGTCTGCTCATCTGAGCTCAGCTGGCTCACACAGCTCTAGGGGCTG 1860  
1801 CCATGACTTAAAGGTCTGCTCATCTGAGCTCAGCTGGCTCACACAGCTCTAGGGGCTG 1860  
1861 CTCCTCTAACTCAGCTGGGTTTGTGAGGCTCTGTGGCCAGAGAGAGCTCGATATCT 1920  
1861 CTCCTCTAACTCAGCTGGGTTTGTGAGGCTCTGTGGCCAGAGAGAGCTCGATATCT 1920  
1921 GAGCAAAAATAGCAAAAGGCTCTCTCAGGCCACTGGGCTTGAATCTTACACTTGAAGCAAC 1980  
1921 GAGCAAAAATAGCAAAAGGCTCTCTCAGGCCACTGGGCTTGAATCTTACACTTGAAGCAAC 1980  
1981 TTGCTGGCACCCCGCTCCCAACCTTCTTGGCTGGGTAGGAGAGGCTAAAGATCACCC 2040  
1981 TTGCTGGCACCCCGCTCCCAACCTTCTTGGCTGGGTAGGAGAGGCTAAAGATCACCC 2040  
2041 TAAATTTACTCATCTCTAGTCTGCCTCACAATGGGCTCAGAGCTCCCGAGCACCA 2100  
2041 TAAATTTACTCATCTCTAGTCTGCCTCACAATGGGCTCAGAGCTCCCGAGCACCA 2100  
2101 ATTACAGGTCACCCCTCTCTTCTGCACTGTCCCAAACTTGTCTGTCAAATCCGAGATC 2160  
2101 ATTACAGGTCACCCCTCTCTTCTGCACTGTCCCAAACTTGTCTGTCAAATCCGAGATC 2160  
2161 TAAATCTCCCTTACGCTCTGCGAGGAATTTCTTTCAGACCTCACTAGGCAAGCCGGTTG 2220  
2161 TAAATCTCCCTTACGCTCTGCGAGGAATTTCTTTCAGACCTCACTAGGCAAGCCGGTTG 2220  
2221 CTCCTTGTGAGGAGAAATTTGTAGTCAATTTCTCACTTCAAAATCTTGGGGCTGATCTTCT 2280  
2221 CTCCTTGTGAGGAGAAATTTGTAGTCAATTTCTCACTTCAAAATCTTGGGGCTGATCTTCT 2280  
2281 CTCACTTGTGACCCCAACCTCTGTAAATAGATTTACCGCAATTTACGGCTGCATTTCTGTA 2340  
2281 CTCACTTGTGACCCCAACCTCTGTAAATAGATTTACCGCAATTTACGGCTGCATTTCTGTA 2340  
2341 GTGGGATGCTCTTCTTAAATGGAGGAGTGTTCATTTATTAATAGTTATTACCTGAGTAT 2400  
2341 GTGGGATGCTCTTCTTAAATGGAGGAGTGTTCATTTATTAATAGTTATTACCTGAGTAT 2400  
2401 GCAATAAAGATGTGGTGGCCACTCTTTTCTGTTGGTGGGAGCA 2443  
2401 GCAATAAAGATGTGGTGGCCACTCTTTTCTGTTGGTGGGAGCA 2443



1141 GAACAGAGCCTACTCTCAAGAGAAATCACTCAAGGTTTTTGAAGAGACAGGGACACGCT 1200  
1201 CTATGCCCCCTATTCACACATTTTTCAGCTGCGAGACCCCTCCCAAGAAAGAAATTC 1260  
1201 CTATGCCCCCTATTCACACATTTTTCAGCTGCGAGACCCCTCCCAAGAAAGAAATTC 1260  
1261 CATCCCAAGGGCCACGCTTGGCGGAGCCCTTACAAAGACTATGAAGTAAAGAAAGAGG 1320  
1261 CATCCCAAGGGCCACGCTTGGCGGAGCCCTTACAAAGACTATGAAGTAAAGAAAGAGG 1320  
1321 CAGCTAACTCTGTCTGAAGAGTGGGACAAATCAGCCGCGCGAGATCTAGCGGAGC 1380  
1321 CAGCTAACTCTGTCTGAAGAGTGGGACAAATCAGCCGCGCGAGATCTAGCGGAGC 1380  
1381 TCAAGGGATGTGGGGAATCTTGAGTCTTCTGAGAAACTGTACAGACACTACGGGA 1440  
1381 TCAAGGGATGTGGGGAATCTTGAGTCTTCTGAGAAACTGTACAGACACTACGGGA 1440  
1441 ACAGTTTGCCTCCCTCCAGCCTCAACCAAAATCTTCCATGCTGGGGCTGATGTGGCT 1500  
1441 ACAGTTTGCCTCCCTCCAGCCTCAACCAAAATCTTCCATGCTGGGGCTGATGTGGCT 1500  
1501 AGTAAGACTCCAGTTCTTAGAGCGCTGTAGTATTTTTTTTTTTTGTCTCATCTTTGTT 1560  
1501 AGTAAGACTCCAGTTCTTAGAGCGCTGTAGTATTTTTTTTTTTTGTCTCATCTTTGTT 1560  
1561 ATACTTTCTTTTAAGTGGGAGTCTCAGGCAACTCAAGTTTAGACCCCTTACTCTTTGTT 1620  
1561 ATACTTTCTTTTAAGTGGGAGTCTCAGGCAACTCAAGTTTAGACCCCTTACTCTTTGTT 1620  
1621 GTTTTTTGAACAGGATCTTGCTCTGTCAACCAAGCTTGTAGTCAGTGGTGCATCACAG 1680  
1621 GTTTTTTGAACAGGATCTTGCTCTGTCAACCAAGCTTGTAGTCAGTGGTGCATCACAG 1680  
1681 CCAGTGCAGCCTCGACCACTGTGTCTGAAGCAATCTCCATCTCCCAAGTG 1740  
1681 CCAGTGCAGCCTCGACCACTGTGTCTGAAGCAATCTCCATCTCCCAAGTG 1740  
1741 CTGGGATGACAGCGTGAAGCAAGCTCCAGCCTAGGCGCTTAATCTTGTCTTTT 1800  
1741 CTGGGATGACAGCGTGAAGCAAGCTCCAGCCTAGGCGCTTAATCTTGTCTTTT 1800  
1801 CCATGACTAAAGTCTGTGCTCATCTGAGCTCAGCTGGCTCAGACAGCTTAGGGCCCTG 1860  
1801 GAGCAAAATAGAGTCTGTGCTCATCTGAGCTCAGCTGGCTCAGACAGCTTAGGGCCCTG 1860  
1861 CTCCTCTAACTCAGTGGGTTTTGTGAGCTCTGTGGCCAGAGCAGACCTGCATATCT 1920  
1861 CTCCTCTAACTCAGTGGGTTTTGTGAGCTCTGTGGCCAGAGCAGACCTGCATATCT 1920  
1921 GAGCAAAATAGCAAAAGCCTCTCAGCCCACTGGCCTGAATCTACACTGGAAGCAAC 1980  
1921 GAGCAAAATAGCAAAAGCCTCTCAGCCCACTGGCCTGAATCTACACTGGAAGCAAC 1980  
1981 TTCTGGCACCCCGCTCCCAACCTCTTGTGCTGGTAGGAGAGCTAAAGATCACCC 2040  
1981 TTCTGGCACCCCGCTCCCAACCTCTTGTGCTGGTAGGAGAGCTAAAGATCACCC 2040  
2041 TAAATTTACTCATCTCTTAGTGTCTCCTCAGCTTGGGCTCAGAGCTCCCAAGCAAC 2100  
2041 TAAATTTACTCATCTCTTAGTGTCTCCTCAGCTTGGGCTCAGAGCTCCCAAGCAAC 2100  
2101 ATTCAGGTCACCCCTCTCTTGTGACTGTCCCAACTTGTCTCAATTCGAGATC 2160  
2101 ATTCAGGTCACCCCTCTCTTGTGACTGTCCCAACTTGTCTCAATTCGAGATC 2160  
2161 TAATCTCCCTCAGCTCTGCGAGGAATCTTTTTCAGACCTCTCAGTACAGAGCCCGGTTG 2220  
2161 TAATCTCCCTCAGCTCTGCGAGGAATCTTTTTCAGACCTCTCAGTACAGAGCCCGGTTG 2220  
2221 CTCCTTTGTCAGGAAATTTGTAGATCATTTCTCAATCTTCAATCTTGGGCTGATCTTCT 2280  
2221 CTCCTTTGTCAGGAAATTTGTAGATCATTTCTCAATCTTCAATCTTGGGCTGATCTTCT 2280

QY 2281 CTCATCTTGCACCCCAACCTCTGTAATAGATTACGGATTTCGGCTGCATTTCTGTAA 2340  
Db 2281 CTCATCTTGCACCCCAACCTCTGTAATAGATTACGGATTTCGGCTGCATTTCTGTAA 2340  
QY 2341 GTGGGCATGCTCTCTTAATGGAGAGTGTTCATTGTATATAAGTTATTTCACCTGAGTAT 2400  
Db 2341 GTGGGCATGCTCTCTTAATGGAGAGTGTTCATTGTATATAAGTTATTTCACCTGAGTAT 2400  
QY 2401 GCATTAAGATGTGTGGGCACCTCTTTCATGTGTGGTGGCAGCAAAAAA 2456  
Db 2401 GCATTAAGATGTGTGGGCACCTCTTTCATGTGTGGTGGCAGCAAAAAA 2456

## RESULT 14

AAH14688  
ID AAH14688 standard; cDNA; 2446 BP.

XX AAH14688;

DT 26-JUN-2001 (first entry)

XX Human cDNA sequence SEQ ID NO:12388.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-00116126.

XX 29-JUL-1999; 99JP-00248036.

XX 27-AUG-1999; 99JP-00300253.

XX 11-JAN-2000; 2000JP-00118776.

XX 02-MAY-2000; 2000JP-00183767.

XX 09-JUN-2000; 2000JP-00241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

DR Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

XX Claim 8; SEQ ID NO 12388; 2537pp + Sequence Listing; English.

XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the 5602 complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893

CC represent human amino acid sequences; and AAH13629 to AAH13632 represent CC oligonucleotides, all of which are used in the exemplification of the present invention

XX Sequence 2446 BP; 507 A; 723 C; 572 G; 644 T; 0 U; 0 Other;

SQ Query Match 99.3%; Score 2439.8; DB 4; Length 2446;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 2441; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY	1	ATAACAGCATGAAGTCGCTGGAACTGGAAATAGGCGTGCCTCTCCGACCCCTCCCC	60
DB	1	ATAACAGCATGAAGTCGCTGGAACTGGAAATAGGCGTGCCTCTCCGACCCCTCCCC	60
QY	61	TCCTTGCTCCCTGTGCTCACCCCTCGCTCGTTCCTCCCTCCGCGAGGCGCGCTTTATA	120
DB	61	TCCTTGCTCCCTGTGCTCACCCCTCGCTCGTTCCTCCCTCCGCGAGGCGCGCTTTATA	120
QY	121	ACAACTGCTCAGAGTCGAGGGGGGATAGTGTCAAGGTCTCCCCAGCACTAGAGAG	180
DB	121	ACAACTGCTCAGAGTCGAGGGGGGATAGTGTCAAGGTCTCCCCAGCACTAGAGAG	180
QY	181	CTCGCTGCTGCCCTCTTCGCGCGGGAGAGCAGCACCAAGTTACGGCCCAACGCTTTGGC	240
DB	181	CTCGCTGCTGCCCTCTTCGCGCGGGAGAGCAGCACCAAGTTACGGCCCAACGCTTTGGC	240
QY	241	ACTAGGGTCCAGAAATGGCTTACACAGTCCCTGATGTTGCGCAATGGCTGAAATCCAA	300
DB	241	ACTAGGGTCCAGAAATGGCTTACACAGTCCCTGATGTTGCGCAATGGCTGAAATCCAA	300
QY	301	GTACTACAGACTTTGTGATAAGGCTGAAGCTTGGGGCATCGCTAGAAACGGTGGCCAC	360
DB	301	GTACTACAGACTTTGTGATAAGGCTGAAGCTTGGGGCATCGCTAGAAACGGTGGCCAC	360
QY	361	AGCGGGGTGTGACCTCGCTGGCTTATGCTCACTCTCCGATCCTCGTCTGCAAGGT	420
DB	361	AGCGGGGTGTGACCTCGCTGGCTTATGCTCACTCTCCGATCCTCGTCTGCAAGGT	420
QY	421	GCAGGACTCCAAACAGCGGAAATGCTGCTACTCAAGTTTCTCTCTCTCGGTGTGTT	480
DB	421	GCAGGACTCCAAACAGCGGAAATGCTGCTACTCAAGTTTCTCTCTCTCGGTGTGTT	480
QY	481	GGGCATCTTTGGGCTCACCTTGCCTTCATCGGACTGGAGGAGACAGGGGCCAC	540
DB	481	GGGCATCTTTGGGCTCACCTTGCCTTCATCGGACTGGAGGAGACAGGGGCCAC	540
QY	541	ACGCTTCTTCTCTTTGGGATCCTCTTTTCCATCTGCTCTCTGCTGCTGCTCATGC	600
DB	541	ACGCTTCTTCTCTTTGGGATCCTCTTTTCCATCTGCTCTCTGCTGCTGCTCATGC	600
QY	601	TGTCAAGTCTGACCAAGCTCGTCGGGGGAGGAGCCCTTTCCCTGTTGGTGAATCTGGG	660
DB	601	TGTCAAGTCTGACCAAGCTCGTCGGGGGAGGAGCCCTTTCCCTGTTGGTGAATCTGGG	660
QY	661	TCGCGCGTGGGCTTCAGCCTAGTCCAGATGTTATCGCTATTTGAATATTTGCTCTGAC	720
DB	661	TCGCGCGTGGGCTTCAGCCTAGTCCAGATGTTATCGCTATTTGAATATTTGCTCTGAC	720
QY	721	CATGAATAGACCAACGCTCAATGCTCTTTTCTGAGCTTCCGCTCTCTGCTGCAATGAAG	780
DB	721	CATGAATAGACCAACGCTCAATGCTCTTTTCTGAGCTTTCGCTCTCTGCTGCAATGAAG	780
QY	781	CTTTGTCCTCTGCTCACCTACGTCCTCTTTTGAATGGCGCTGACCTTCTCATGTCCTC	840
DB	781	CTTTGTCCTCTGCTCACCTACGTCCTCTTTTGAATGGCGCTGACCTTCTCATGTCCTC	840
QY	841	CTTACCTTCTGTGGTTCCTTACGGGCTGGAGAGACATAGGGGCCACATCTACCTCAC	900
DB	841	CTTACCTTCTGTGGTTCCTTACGGGCTGGAGAGACATAGGGGCCACATCTACCTCAC	900
QY	901	GATGCTCTCTCCATTTGCCATCTGGGTGGCTGGATCACCCTGCTCATGCTTCTGACTT	960
DB	901	GATGCTCTCTCCATTTGCCATCTGGGTGGCTGGATCACCCTGCTCATGCTTCTGACTT	960



Db 841 CTTACCTTCTGTGTTCTCTTACGGGCTGGAGAGACATGSGGCCACATCTACCTCAC 900  
Qy 901 GATGCTCCTCTCCATGTCOATCTGGGTGGCCTGGATCACCTCTCATGCTTCTCTGACTT 960  
Db 901 GATGCTCCTCTCCATGTCOATCTGGGTGGCCTGGATCACCTCTCATGCTTCTCTGACTT 960  
Qy 961 TGACCGCAGGTGGGATGACACCATCTCTAGCTCCGCTTGGCTGGCAATGGCTGGGTGTT 1020  
Db 961 TGACCGCAGGTGGGATGACACCATCTCTAGCTCCGCTTGGCTGGCAATGGCTGGGTGTT 1020  
Qy 1021 CCTGTTGGCTTATGTTAGTCCGAGTTTGGCTGCTCACAAGCAACGAAACCCCATGGA 1080  
Db 1021 CCTGTTGGCTTATGTTAGTCCGAGTTTGGCTGCTCACAAGCAACGAAACCCCATGGA 1080  
Qy 1081 TTATCCTGTTAGGATGCTTCTGTAACCTCAACTCGTGAAGAGAGCTATGGTGTGGA 1140  
Db 1081 TTATCCTGTTAGGATGCTTCTGTAACCTCAACTCGTGAAGAGAGCTATGGTGTGGA 1140  
Qy 1141 GAACAGAGCTACTCTCAAGAGGAAATCACTCAAGGTTTGAAGAGACAGGCGACCGCT 1200  
Db 1141 GAACAGAGCTACTCTCAAGAGGAAATCACTCAAGGTTTGAAGAGACAGGCGACCGCT 1200  
Qy 1201 CTATGCCCTTATTCACACATTTTCAGCTGAGAACAGCGCTCCCCAAAGAAATTTCTC 1260  
Db 1201 CTATGCCCTTATTCACACATTTTCAGCTGAGAACAGCGCTCCCCAAAGAAATTTCTC 1260  
Qy 1261 CATCCACGGGCCACCGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAGAAAGAGGG 1320  
Db 1261 CATCCACGGGCCACCGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAGAAAGAGGG 1320  
Qy 1321 CAGCTAACTCTGTCTGAAGAGTGGGACAAATGCAGCCGGCGGCAGATCTAGCGGGAGC 1380  
Db 1321 CAGCTAACTCTGTCTGAAGAGTGGGACAAATGCAGCCGGCGGCAGATCTAGCGGGAGC 1380  
Qy 1381 TCAAGAGGATGTGGGCGAAATCTTGAGTCTTCTGAGAAACTGTACAAGACACTACGGGA 1440  
Db 1381 TCAAGAGGATGTGGGCGAAATCTTGAGTCTTCTGAGAAACTGTACAAGACACTACGGGA 1440  
Qy 1441 ACAGTTTGCTCCCTCCAGCCTCAACCAATTTCTTCATGCTGGGGCTGATGGGCT 1500  
Db 1441 ACAGTTTGCTCCCTCCAGCCTCAACCAATTTCTTCATGCTGGGGCTGATGGGCT 1500  
Qy 1501 AGTAAGACTCCAGTCTTAGAGCGCTGTAGTATTTTTTTTTTTTTTTTGTCTCATCCTTTGG 1560  
Db 1501 AGTAAGACTCCAGTCTTAGAGCGCTGTAGTATTTTTTTTTTTTTTTTGTCTCATCCTTTGG 1560  
Qy 1561 ATACTTCTTTTTAAGTGGGAGTCTCAGGCAACTCAAGTTTACAGCCCTTACTCTTTTGTGTT 1620  
Db 1561 ATACTTCTTTTTAAGTGGGAGTCTCAGGCAACTCAAGTTTACAGCCCTTACTCTTTTGTGTT 1620  
Qy 1621 GTTTTTTGAAACAGGATCTTGCTGTCTACCCAGGCTTGAGTGCAGTGGTGGATCACAG 1680  
Db 1621 GTTTTTTGAAACAGGATCTTGCTGTCTACCCAGGCTTGAGTGCAGTGGTGGATCACAG 1680  
Qy 1681 CCCAGTGCAGCTCGACCACTGTCTCAAGCAATCTCCCATCTCCATCTCCCAAGTG 1740  
Db 1681 CCCAGTGCAGCTCGACCACTGTCTCAAGCAATCTCCCATCTCCATCTCCCAAGTG 1740  
Qy 1741 CTGGGATGACAGCGTGAGCCACAGCTCCAGCCTAGGCGCTTAATCTTTGCTGTTATTTT 1800  
Db 1741 CTGGGATGACAGCGTGAGCCACAGCTCCAGCCTAGGCGCTTAATCTTTGCTGTTATTTT 1800  
Qy 1801 CCATGGACTAAAGGTCTGGTCACTCGAGCTACGCTGGCTCAACAGCTCTAGGGCCCTG 1860  
Db 1801 CCATGGACTAAAGGTCTGGTCACTCGAGCTACGCTGGCTCAACAGCTCTAGGGCCCTG 1860  
Qy 1861 CTCCTCTAACTACAGTGGGTTTTGTGAGGCTGTGGCCAGAGCAGACCTGCATATCT 1920  
Db 1861 CTCCTCTAACTACAGTGGGTTTTGTGAGGCTGTGGCCAGAGCAGACCTGCATATCT 1920  
Qy 1921 GAGCAAAAATAGCAAAAGCCTCTCTCAGGCCACTGGCCTGAATCTACACTGGAAGCCAAAC 1980  
Db 1921 GAGCAAAAATAGCAAAAGCCTCTCTCAGGCCACTGGCCTGAATCTACACTGGAAGCCAAAC 1980

Search completed: June 10, 2006, 15:19:12  
Job time : 1410 secs

Qy 1981 TTGCTGGCACCCCGCTCCCCAAACCTTCTTGCCTGGGTAGGAGAGGCTAAAGATCACCC 2040  
Db 1981 TTGCTGGCACCCCGCTCCCCAAACCTTCTTGCCTGGGTAGGAGAGGCTAAAGATCACCC 2040  
Qy 2041 TAAATTTTACTCATCTCTCTAGTGTGCTCACAATTTGGGCTCAGCAGCTCCCCAGCACCA 2100  
Db 2041 TAAATTTTACTCATCTCTCTAGTGTGCTCACAATTTGGGCTCAGCAGCTCCCCAGCACCA 2100  
Qy 2101 ATTTCAGGTCAACCCCTCTCTTCTTGGCACTGTCCCCAAACTTTGCTGTCAATTTCCGAGATC 2160  
Db 2101 ATTTCAGGTCAACCCCTCTCTTCTTGGCACTGTCCCCAAACTTTGCTGTCAATTTCCGAGATC 2160  
Qy 2161 TAAATCTCCCGCTACGCTCTGCCAGGAAATTTCTTTTCAGACCTCACTAGCACAAAGCCCGTTG 2220  
Db 2161 TAAATCTCCCGCTACGCTCTGCCAGGAAATTTCTTTTCAGACCTCACTAGCACAAAGCCCGTTG 2220  
Qy 2221 CTCCTTGTGAGAGAAATTTGTAGATCATTTCTCACTTCAAAATTCCTGGGCTGATCTTCT 2280  
Db 2221 CTCCTTGTGAGAGAAATTTGTAGATCATTTCTCACTTCAAAATTCCTGGGCTGATCTTCT 2280  
Qy 2281 CTCATCTTGCACCCCAACCTCTGTAAATAGATTTTACCGCAATTTACGGCTGCATTTCTGTAA 2340  
Db 2281 CTCATCTTGCACCCCAACCTCTGTAAATAGATTTTACCGCAATTTACGGCTGCATTTCTGTAA 2340  
Qy 2341 GTGGGATGCTCTCCTAATGGAGGAGTGTTCATTTGTAATTAATTAATTAATTAATTAATTAAT 2400  
Db 2341 GTGGGATGCTCTCCTAATGGAGGAGTGTTCATTTGTAATTAATTAATTAATTAATTAATTAAT 2400  
Qy 2401 GCAATAAAGATGTGGTGCCACTCTTTTCATGTGTGGCAGCAAAAAA 2456  
Db 2401 GCAATAAAGATGTGGTGCCACTCTTTTCATGTGTGGCAGCAAAAAA 2456

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2006, 10:23:50 ; Search time 200 seconds  
(without alignments)  
816.131 Million cell updates/sec

Title: US-10-600-816-3

Perfect score: 1865

Sequence: 1 MATTVPDGGCRLGKSKYRL.....PRAHWPSPYKDYEYKKEGS 357

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq 8:\*

1: Geneseq1980s:\*

2: Geneseq1990s:\*

3: Geneseq2000s:\*

4: Geneseq2001s:\*

5: Geneseq2002s:\*

6: Geneseq2003as:\*

7: Geneseq2003bs:\*

8: Geneseq2004s:\*

9: Geneseq2005s:\*

10: Geneseq2006s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1865	100.0	357	4	AAB68891 Human REC
2	1865	100.0	357	4	AAB93311 Human pro
3	1865	100.0	357	6	ABR58569 Human can
4	1865	100.0	357	6	ABJ37054 Human bre
5	1865	100.0	357	6	ABR42649 Human GPC
6	1865	100.0	357	6	ABP81984 Human G
7	1865	100.0	357	7	ADD93239 RAIG1. 1/
8	1865	100.0	357	7	ADN39302 Cancer/an
9	1865	100.0	357	8	ADI28460 Human GPC
10	1865	100.0	357	8	ADO29632 Human GPC
11	1865	100.0	357	8	ADR48222 Human ret
12	1865	100.0	357	8	ABM81354 Tumour-as
13	1865	100.0	357	8	ADU06613 Novel bro
14	1865	100.0	357	9	AEA00166 Human TAT
15	1865	100.0	357	9	AEA00190 Human TAT
16	1865	100.0	357	9	AEA00686 Human TAT
17	1865	100.0	357	9	AEA00710 Human TAT
18	1865	100.0	357	9	AED47479 Retinoic
19	1865	100.0	401	8	ADX97565 Pancreati
20	1865	100.0	409	5	ABP41877 Human ova
21	1865	100.0	595	7	ABF70471 Orphan re
22	1861	99.8	357	8	ADI28536 Human GPC
23	1861	99.8	357	8	ADI28535 Human GPC

ALIGNMENTS

RESULT 1

AAB68891

ID AAB68891 standard; protein; 357 AA.

XX

AC AAB68891;

XX

DT 24-APR-2001 (first entry)

XX

DE Human RECAP polypeptide, SEQ ID NO: 21.

XX

Human; RECAP; receptors and associated proteins; cerebroprotective; neurotropic; neuroprotective; anticonvulsant; antiparkinsonian; anti-HIV; antidiabetic; immunostimulant; immunomodulator; antiinflammatory; antithyroid; immunosuppressive; nephrotropic; antigout; thyromimetic; cytostatic; antibacterial; virucide; fungicide; protozoacide; antiarteriosclerotic; hepatotropic; gene therapy; infection; cancer.

OS Homo sapiens.

XX

PN WO200107612-A2.

XX

PD 01-FEB-2001.

XX

PF 21-JUL-2000; 2000WO-US020035.

XX

PR 21-JUL-1999; 99US-0145232P.

PR

07-OCT-1999; 99US-0158578P.

PR

12-NOV-1999; 99US-0165192P.

XX

PA (INCY-) INCYTE GENOMICS INC.

XX

PI Au-Young J, Bandman O, Tang YT, Yue H, Azimzai Y, Burford N;

PI

Baughn MR, Lu DAM, Hillman JL, Patterson C, Lal P;

XX

DR WPI: 2001-168554/17.

DR

N-PSDB; AAF58615.

XX

Novel receptors and associated proteins for diagnosis and treatment of neurological disorders, immunological disorders including autoimmune/inflammatory disorders and cell proliferative disorders such as cancer.

PS Disclosure; Page 111-112; 128pp; English.

XX

The present sequence is a human RECAP (receptors and associated proteins) polypeptide. RECAP polynucleotides and polypeptides are useful in the diagnosis, treatment and prevention of neurological disorders such as



CC stroke, Alzheimer's disease, Pick's disease, Huntington's disease,  
 CC dementia, Parkinson's disease, Down's syndrome, amyotrophic lateral  
 CC sclerosis, multiple sclerosis, bacterial and viral meningitis, CJD  
 CC (Creutzfeldt-Jakob disease), GSS (Gerstmann-Strausler-Scheinker  
 CC syndrome); immunological disorders, including autoimmune/inflammatory  
 CC disorders such as AIDS, Digeorge's syndrome, severe combined  
 CC immunodeficiency disease (SCID), Chediak-Higashi syndrome, Cushing's  
 CC disease, Addison's disease, autoimmune thyroiditis, Crohn's disease,  
 CC diabetes mellitus, Good pasture's syndrome, gout, Grave's disease,  
 CC Hashimoto's thyroiditis, Sjogren's syndrome, Werner's syndrome, viral,  
 CC bacterial, fungal, parasitic, protozoal, and helminthic infections; and  
 CC cell proliferation disorders such as arteriosclerosis, atherosclerosis,  
 CC cirrhosis, hepatitis and cancer  
 XX  
 XX Sequence 357 AA;

Query Match 100.0%; Score 1865; DB 4; Length 357;  
 Best Local Similarity 100.0%; Pred. No. 2.7e-203; Indels 0; Gaps 0;  
 Matches 357; Conservative 0; Mismatches 0;

QY 1 MATTPDGCGRNGLSKYYRLCDKAEAWGIVLFTVATAGVWTSVAFMLTLPILVCKVQDSN 60  
 DB 1 MATTPDGCGRNGLSKYYRLCDKAEAWGIVLFTVATAGVWTSVAFMLTLPILVCKVQDSN 60

QY 61 RKMLPTQFLFLGLVGLGIFGLTFAFIIGLDGSTGPTFRFLFGILFSCFSCLLAHAVSLT 120  
 DB 61 RKMLPTQFLFLGLVGLGIFGLTFAFIIGLDGSTGPTFRFLFGILFSCFSCLLAHAVSLT 120

QY 121 KLVGRKPLSLVILGLAVGSLVQDVTAIEYIVLTMNTNNTNVSSELSAPRNEDFVLL 180  
 DB 121 KLVGRKPLSLVILGLAVGSLVQDVTAIEYIVLTMNTNNTNVSSELSAPRNEDFVLL 180

QY 181 LTYVFLMALTFELMSSTFCGSFTGKRGHAIYLTMLLSIAIWAWITLLMLPDPDRRW 240  
 DB 181 LTYVFLMALTFELMSSTFCGSFTGKRGHAIYLTMLLSIAIWAWITLLMLPDPDRRW 240

QY 241 DDTILSSALAANGWVFLLAYVSPFELWLTQRNPMDYPVEDAFCKPQLVKKSYGVENRAY 300  
 DB 241 DDTILSSALAANGWVFLLAYVSPFELWLTQRNPMDYPVEDAFCKPQLVKKSYGVENRAY 300

QY 301 SQEETITQGFETGDTLYAPYSTHFOQNQPPQKFSIPRAHAWPSYKDYEVKKEGS 357  
 DB 301 SQEETITQGFETGDTLYAPYSTHFOQNQPPQKFSIPRAHAWPSYKDYEVKKEGS 357

RESULT 2  
 AAB93311  
 ID AAB93311 standard; protein; 357 AA.  
 XX  
 AC AAB93311;  
 XX  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Human protein sequence SEQ ID NO:12389.  
 XX  
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1074617-A2.  
 XX  
 XX 07-FEB-2001.  
 PD  
 XX 28-JUL-2000; 2000EP-00116126.  
 PF  
 XX 29-JUL-1999; 99JP-00248036.  
 PR 27-AUG-1999; 99JP-00300253.  
 PR 11-JAN-2000; 2000JP-00118776.  
 PR 02-MAY-2000; 2000JP-00183767.  
 PR 09-JUN-2000; 2000JP-00241899.  
 XX  
 XX (HELI-) HELIX RES INST.

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 DR WPI; 2001-318749/34.  
 XX  
 XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-  
 PT length cDNAs defined in the specification, and for the detection and/or  
 PT diagnosis of the abnormality of the proteins encoded by the full-length  
 XX cDNAs.  
 PS Claim 8; SEQ ID NO 12389; 2537pp + Sequence Listing; English.  
 XX  
 CC The present invention describes primer sets for synthesizing 5602 full-  
 CC length cDNAs defined in the specification. Where a primer set comprises:  
 CC (a) an oligo-dT primer and an oligonucleotide complementary to the  
 CC complementary strand of a polynucleotide which comprises one of the 5602  
 CC nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC polynucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the  
 CC specification. The primer sets can be used in antisense therapy and in  
 CC gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893  
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent  
 CC oligonucleotides, all of which are used in the exemplification of the  
 CC present invention  
 XX  
 XX Sequence 357 AA;

Query Match 100.0%; Score 1865; DB 4; Length 357;  
 Best Local Similarity 100.0%; Pred. No. 2.7e-203;  
 Matches 357; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATTPDGCGRNGLSKYYRLCDKAEAWGIVLFTVATAGVWTSVAFMLTLPILVCKVQDSN 60  
 DB 1 MATTPDGCGRNGLSKYYRLCDKAEAWGIVLFTVATAGVWTSVAFMLTLPILVCKVQDSN 60

QY 61 RKMLPTQFLFLGLVGLGIFGLTFAFIIGLDGSTGPTFRFLFGILFSCFSCLLAHAVSLT 120  
 DB 61 RKMLPTQFLFLGLVGLGIFGLTFAFIIGLDGSTGPTFRFLFGILFSCFSCLLAHAVSLT 120

QY 121 KLVGRKPLSLVILGLAVGSLVQDVTAIEYIVLTMNTNNTNVSSELSAPRNEDFVLL 180  
 DB 121 KLVGRKPLSLVILGLAVGSLVQDVTAIEYIVLTMNTNNTNVSSELSAPRNEDFVLL 180

QY 181 LTYVFLMALTFELMSSTFCGSFTGKRGHAIYLTMLLSIAIWAWITLLMLPDPDRRW 240  
 DB 181 LTYVFLMALTFELMSSTFCGSFTGKRGHAIYLTMLLSIAIWAWITLLMLPDPDRRW 240

QY 241 DDTILSSALAANGWVFLLAYVSPFELWLTQRNPMDYPVEDAFCKPQLVKKSYGVENRAY 300  
 DB 241 DDTILSSALAANGWVFLLAYVSPFELWLTQRNPMDYPVEDAFCKPQLVKKSYGVENRAY 300

QY 301 SQEETITQGFETGDTLYAPYSTHFOQNQPPQKFSIPRAHAWPSYKDYEVKKEGS 357  
 DB 301 SQEETITQGFETGDTLYAPYSTHFOQNQPPQKFSIPRAHAWPSYKDYEVKKEGS 357

RESULT 3  
 ABR58569  
 ID ABR58569 standard; protein; 357 AA.  
 XX  
 AC ABR58569;  
 XX  
 DT 09-JUL-2003 (first entry)





```
QY 181 LTYVLFLMALTFMSSFTFCGSFTGKRGHAIYLTMLLSIAIWAVITLMLPDDRRW 240
DB 181 LTYVLFLMALTFMSSFTFCGSFTGKRGHAIYLTMLLSIAIWAVITLMLPDDRRW 240
QY 241 DDTILSSALAANGWVFLAYVSPFWLLTKQNPMDYPVEDAFCKPOLVKKSYGVENRAY 300
DB 241 DDTILSSALAANGWVFLAYVSPFWLLTKQNPMDYPVEDAFCKPOLVKKSYGVENRAY 300
QY 301 SQEETITQGFEEETGDTLYAPYSTHFQLNQPPQKEFSIPRAHAWPSPYKDYEVKKSGS 357
DB 301 SQEETITQGFEEETGDTLYAPYSTHFQLNQPPQKEFSIPRAHAWPSPYKDYEVKKSGS 357

RESULT 5
ID ABR42649 standard; protein; 357 AA.
XX
AC ABR42649;
XX
DT 26-AUG-2003 (first entry)
XX
DE Human GPCR-like retinoic acid-induced gene 1 protein.
XX
KW Human; retinoic acid-induced gene 1; RAIG1; feeding; fasting; GPCR;
KW receptor; G-protein coupled receptor; anorectic; antidiabetic;
KW antidepressant; immunomodulator; transgenic; gene therapy.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Modified-site 4..8 /note= "phosphorylated by casein kinase II"
FT Modified-site 8..14 /note= "N-myristoylated"
FT Modified-site 38..43 /note= "N-myristoylated"
FT Modified-site 59..61 /note= "phosphorylated by protein kinase C"
FT Modified-site 80..86 /note= "N-myristoylated"
FT Modified-site 88..93 /note= "N-myristoylated"
FT Modified-site 102..107 /note= "N-myristoylated"
FT Modified-site 124..127 /note= "Amidated"
FT Modified-site 136..142 /note= "N-myristoylated"
FT Modified-site 158..161 /note= "N-glycosylated"
FT Modified-site 201..206 /note= "N-myristoylated"
FT Modified-site 301..304 /note= "phosphorylated by casein kinase II"
FT
XX
PN WO2003016553-A2.
XX
PD 27-FEB-2003.
XX
PF 20-AUG-2002; 2002WO-US026510.
XX
PR 20-AUG-2001; 2001US-0313940P.
XX
PA (GETH ) GENENTECH INC.
PA (CURA-) CURAGEN CORP.
XX
PI Lewin DA, Stewart TA;
XX
DR WPI; 2003-278580/27.
DR N-PSDB; ACC58386.
XX
, PT New G-protein coupled receptor-like retinoic acid induced gene 1 (GPCR-
```

```
PT like RAIG1) polypeptide and gene, useful for diagnosing or treating
PT metabolic disorders, e.g. obesity, anorexia, cachexia or diabetes.
XX
PS Disclosure; Page 19-20; 150pp; English.
XX
CC The present sequence is that of human G-protein coupled receptor-like
CC retinoic acid induced gene 1 (GPCR-like RAIG1) protein. This is the human
CC homologue of murine GPCR-like RAIG1 (see ABR42648). The murine GPCR-like
CC RAIG1 gene was shown to be differentially regulated during fasting-
CC feeding cycles, with moderate induction early in fasting, down-regulation
CC with extended fasting and 4-fold up-regulation with feeding in recovery
CC from fasting. The differentially expressed gene, its mRNA, and the
CC encoded protein, can each be manipulated to detect and treat metabolic
CC disorders associated with up- or down-regulation of GPCR-like RAIG1
CC activity, such as obesity, anorexia, cachexia or diabetes
XX
SQ Sequence 357 AA;

Query Match 100.0%; Score 1865; DB 6; Length 357;
Best Local Similarity 100.0%; Pred. No. 2.7e-203;
Matches 357; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATTVPGCRNGLSKSYRLCDKAEAWGIVLETATAGVWTSVAFMLTLPILVCKVQDSN 60
DB 1 MATTVPGCRNGLSKSYRLCDKAEAWGIVLETATAGVWTSVAFMLTLPILVCKVQDSN 60
QY 61 RRKMLPTQFLFLGLVGLGIFGLTFAFIIGLDGSTGPTFRFLFGILFSCFCLLAHAVSLT 120
DB 61 RRKMLPTQFLFLGLVGLGIFGLTFAFIIGLDGSTGPTFRFLFGILFSCFCLLAHAVSLT 120
QY 121 KLVGRKPLSLVILGLAVGSLVQDVIAIEYIVLTMRNTNVNVSSEISAPRNEDFVLL 180
DB 121 KLVGRKPLSLVILGLAVGSLVQDVIAIEYIVLTMRNTNVNVSSEISAPRNEDFVLL 180
QY 181 LTYVLFLMALTFMSSFTFCGSFTGKRGHAIYLTMLLSIAIWAVITLMLPDDRRW 240
DB 181 LTYVLFLMALTFMSSFTFCGSFTGKRGHAIYLTMLLSIAIWAVITLMLPDDRRW 240
QY 241 DDTILSSALAANGWVFLAYVSPFWLLTKQNPMDYPVEDAFCKPOLVKKSYGVENRAY 300
DB 241 DDTILSSALAANGWVFLAYVSPFWLLTKQNPMDYPVEDAFCKPOLVKKSYGVENRAY 300
QY 301 SQEETITQGFEEETGDTLYAPYSTHFQLNQPPQKEFSIPRAHAWPSPYKDYEVKKSGS 357
DB 301 SQEETITQGFEEETGDTLYAPYSTHFQLNQPPQKEFSIPRAHAWPSPYKDYEVKKSGS 357

RESULT 6
ABP81984
ID ABP81984 standard; protein; 357 AA.
XX
AC ABP81984;
XX
DT 04-MAR-2003 (first entry)
XX
DE Human G protein-coupled receptor RAIG1 protein SEQ ID NO:454.
XX
KW G protein-coupled receptor; GPCR; antigenic peptide; gene therapy;
KW G protein-coupled receptor modulator; antibody; immune-related disease;
KW growth-related disease; cell regeneration-related disease; AIDS; cancer;
KW immunological-related cell proliferative disease; autoimmune disease;
KW Alzheimer's disease; atherosclerosis; infection; osteoarthritis; allergy;
KW osteoporosis; cardiomyopathy; inflammation; Crohn's disease; diabetes;
KW graft versus host disease; Parkinson's disease; multiple sclerosis; pain;
KW psoriasis; anxiety; depression; schizophrenia; dementia; memory loss;
KW mental retardation; epilepsy; asthma; tuberculosis; obesity; nausea;
KW hypertension; hypotension; renal disorder; rheumatoid arthritis; trauma;
KW ulcer.
XX
OS Homo sapiens.
XX
PN WO200261087-A2.
XX
```

08-AUG-2002.  
19-DEC-2001; 2001WO-US050107.  
19-DEC-2000; 2000US-0257144P.  
(LIFE-) LIFESPAN BIOSCIENCES INC.  
Burger GC, Roush CL, Brown JP;  
WPI; 2003-046718/04.  
N-PSDB; AB242832.  
New isolated antigenic peptides e.g., for G protein-coupled receptors (GPCR), useful for diagnosing and designing drugs for treating conditions in which GPCRs are involved, e.g. AIDS, Alzheimer's disease, cancer or autoimmune diseases.  
Disclosure; Fig 1; 523pp; English.  
The present invention describes antigenic peptides (I) comprising: (a) any one of 1601 sequences (see ABP82019 to ABP83619) of 12-24 amino acids. Also described: (1) an assay for the detection of a particular G protein-coupled receptor (GPCR) or a candidate polypeptide in a sample; and (2) an isolated antibody having high specificity and high affinity or avidity for a particular GPCR. (I) can be used as GPCR modulators and in gene therapy. The antigenic peptides for GPCRs are useful in detecting an antibody against a particular GPCR, and in the production of specific antibodies. The peptides and antibodies are also useful for detecting the presence or absence of corresponding GPCRs. The antigenic peptides for GPCRs and antibodies are useful for diagnosing and designing drugs for treating immune-related diseases, growth-related diseases, cell regeneration-related disease, immunological-related cell proliferative diseases, or autoimmune diseases, e.g. AIDS, Alzheimer's disease, atherosclerosis, bacterial, fungal, protozoan or viral infections, osteoarthritis, osteoporosis, cancer, cardiomyopathy, chronic and acute inflammation, allergies, Crohn's disease, diabetes, graft versus host disease, Parkinson's disease, multiple sclerosis, pain, psoriasis, anxiety, depression, schizophrenia, dementia, mental retardation, memory loss, epilepsy, asthma, tuberculosis, obesity, nausea, hypertension, hypotension, renal disorders, rheumatoid arthritis, trauma, ulcers, or any other disorder in which GPCRs are involved. The antibodies may be used in immunoassays and immunodiagnosis. AB242523 to AB242869 encode GPCR proteins given in ABP81675 to ABP82018, which are used in the exemplification of the present invention  
Sequence 357 AA;  
Query Match  
Best Local Similarity 100.0%; Score 1865; DB 6; Length 357;  
Matches 357; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MATTPDGCGRGLKSKYRLCDKAEAGVILETVATAGVVTVAFMLTILPVCKVQDSN 60  
DQ |||||  
1 MATTPDGCGRGLKSKYRLCDKAEAGVILETVATAGVVTVAFMLTILPVCKVQDSN 60  
QY 61 RKMLPTQFLVLGVIGLFAFIIGDSTGPTFFLFGILFISICSCILLAHAVSLT 120  
DQ |||||  
61 RKMLPTQFLVLGVIGLFAFIIGDSTGPTFFLFGILFISICSCILLAHAVSLT 120  
QY 121 KLVGRKPLSLVILGLAVGFSILVQDVIAEYVLTMTNTNNTNVSFELSAPRNEDFVLL 180  
DQ |||||  
121 KLVGRKPLSLVILGLAVGFSILVQDVIAEYVLTMTNTNNTNVSFELSAPRNEDFVLL 180  
QY 181 LTVLFLMALTFMSSFTCGSFTGKRHGAHIYLTMLLSIAIWAIVITLMLPDDRRW 240  
DQ |||||  
181 LTVLFLMALTFMSSFTCGSFTGKRHGAHIYLTMLLSIAIWAIVITLMLPDDRRW 240  
QY 241 DDTILSSALAANGWFLAYVSPFELLTKQNPMDYPVEDAFCKPQLVKKSYGVENRAY 300  
DQ |||||  
241 DDTILSSALAANGWFLAYVSPFELLTKQNPMDYPVEDAFCKPQLVKKSYGVENRAY 300  
QY 301 SBEITQGFETGDTLYAPYSTHFQLONPQPKFSPRAHAWPSPYKDYEVKKEGS 357

Db 301 SBEITQGFETGDTLYAPYSTHFQLONPQPKFSPRAHAWPSPYKDYEVKKEGS 357  
RESULT 7  
ADD93239  
ID ADD93239 standard; protein; 357 AA.  
XX  
AC ADD93239;  
XX  
DT 29-JAN-2004 (first entry)  
XX  
DE RAIG1.  
XX retinoic acid-inducible gene 1; RAIG1; orphan;  
KW G-protein coupled receptor; GPCR; chromosome 12; GPCR5B; carcinoma;  
KW vaccine; breast cancer; pancreatic cancer; lung cancer; liver cancer;  
KW ovarian cancer; colon cancer; osteosarcoma.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Peptide 160..172  
FT Peptide /note="Tandem mass spectrum peptide"  
FT Peptide 334..348  
FT Peptide /note="MALDI mass spectra peptide"  
FT Peptide 340..348  
FT Peptide /note="Tandem mass spectrum peptide"  
XX WO2003087832-A2.  
XX 23-OCT-2003.  
XX 10-APR-2003; 2003WO-GB001587.  
XX 11-APR-2002; 2002GB-00008331.  
PR 17-SEP-2002; 2002GB-00021538.  
XX (OXFO-) OXFORD GLYCOSCIENCES UK LTD.  
XX Terrett JA;  
XX WPI; 2003-845382/78.  
DR N-PSDB; ADD93240.  
XX  
PT Screening, diagnosing and/or treating carcinoma, including breast,  
PT pancreatic, lung, liver, ovarian and colon cancer by detecting the change  
PT in expression or activity of an RAIG1 polypeptide or encoding nucleic  
XX acid molecule.  
PS Claim 1; Fig 1; 43pp; English.  
XX  
CC This sequence represents retinoic acid-inducible gene 1 (RAIG1)  
CC polypeptide. RAIG1 is an orphan G-protein coupled receptor (GPCR) located  
CC on chromosome 12. RAIG1 shows a restricted expression pattern compared to  
CC a related receptor, GPCR5B, which is widely expressed in peripheral and  
CC central tissues. The RAIG1 cDNA and polypeptide sequence may be used in  
CC the method of the invention for screening for and/or diagnosis of carcinoma  
CC carcinoma in a subject, and/or monitoring the effectiveness of carcinoma  
CC therapy. The method comprises detecting and/or quantifying in a  
CC biological sample obtained from the subject an RAIG1 polypeptide and a  
CC nucleic acid molecule. The RAIG1 polypeptide and nucleic acid molecule  
CC are useful in the manufacture of a medicament for the treatment of  
CC carcinoma, where the composition is a vaccine. An agent which interacts  
CC with or causes change in the expression or activity of an RAIG1  
CC polypeptide or nucleic acid molecule, is also useful in the manufacture  
CC of a medicament for the treatment of carcinoma that is breast cancer,  
CC pancreatic cancer, lung cancer, liver cancer, ovarian cancer, colon  
CC cancer and/or osteosarcoma. They can also be used in the diagnosis and  
XX screening of such carcinomas.  
SQ Sequence 357 AA;

Query Match	100.0%;	Score 1865;	DB 7;	Length 357;
Best Local Similarity	100.0%;	Pred. No. 2.7e-203;		
Matches 357;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MATTVPDGCRCNGLSKYYRLCDKAEAWGIVLETVATAGVWTSVAFMLTLPILVCKVQDSN	60	
Db	1	MATTVPDGCRCNGLSKYYRLCDKAEAWGIVLETVATAGVWTSVAFMLTLPILVCKVQDSN	60	
QY	61	RRKMLPTQFLFLLGVLGIFGLTFAPFIIGLDGSTGTRFRLFGILFSCILLAHAVSLT	120	
Db	61	RRKMLPTQFLFLLGVLGIFGLTFAPFIIGLDGSTGTRFRLFGILFSCILLAHAVSLT	120	
QY	121	KLVRGKPLSLVILGLAVGFSLVQDVIAEIVILTMRTNNVFSSELSAPRNEDFVLL	180	
Db	121	KLVRGKPLSLVILGLAVGFSLVQDVIAEIVILTMRTNNVFSSELSAPRNEDFVLL	180	
QY	181	LTYYVLFMLALTFLMSSFTFCGSFTGKRGHAIYLTMLLSIAIWAWITLLMLPDPDRW	240	
Db	181	LTYYVLFMLALTFLMSSFTFCGSFTGKRGHAIYLTMLLSIAIWAWITLLMLPDPDRW	240	
QY	241	DDTILSSALAANGWVFLLAYVSPFELWLTQKRNPDYVEDAFCKPQLVKKSYGVENRAY	300	
Db	241	DDTILSSALAANGWVFLLAYVSPFELWLTQKRNPDYVEDAFCKPQLVKKSYGVENRAY	300	
QY	301	SOEITQGFEEGTDTLYAPYSTHFOLQNPQPKFSPRAHAWPSPYKDYEVKKEGS	357	
Db	301	SOEITQGFEEGTDTLYAPYSTHFOLQNPQPKFSPRAHAWPSPYKDYEVKKEGS	357	
RESULT 8				
ADN39302				
ID	ADN39302	standard; protein; 357 AA.		
XX	AC	ADN39302;		
XX	DT	17-JUN-2004 (first entry)		
XX	DE	Cancer/angiogenesis/fibrosis-related polypeptide, SEQ ID NO:620.		
XX	KW	Human; differential expression; cancer; angiogenic disorder;		
XX	KW	fibrotic disorder; psoriasis; ischaemia; heart disease; atherosclerosis;		
XX	KW	inflammatory disease; autoimmune disease;		
XX	KW	retinal neovascularisation syndrome; scarring; uterine fibroid;		
XX	KW	detection; diagnosis; prognosis; drug screening; drug targeting;		
XX	KW	wound healing; contraception; cytostatic; cardiant; immunomodulatory;		
XX	KW	vulnerary; gene therapy; vaccine.		
XX	OS	Homo sapiens.		
XX	PN	WO2003042661-A2.		
XX	PD	22-MAY-2003.		
XX	PF	13-NOV-2002; 2002WO-US036810.		
XX	PR	13-NOV-2001; 2001US-0350666P.		
XX	PR	21-NOV-2001; 2001US-0332464P.		
XX	PR	29-NOV-2001; 2001US-0334393P.		
XX	PR	03-DEC-2001; 2001US-0335394P.		
XX	PR	14-DEC-2001; 2001US-0340376P.		
XX	PR	08-JAN-2002; 2002US-0347211P.		
XX	PR	10-JAN-2002; 2002US-0347349P.		
XX	PR	13-FEB-2002; 2002US-0355250P.		
XX	PR	20-FEB-2002; 2002US-0359077P.		
XX	PR	29-MAR-2002; 2002US-0368809P.		
XX	PR	04-APR-2002; 2002US-0370110P.		
XX	PR	12-APR-2002; 2002US-0372246P.		
XX	PK	05-JUN-2002; 2002US-0386614P.		
XX	PR	16-JUL-2002; 2002US-0396839P.		
XX	PR	22-JUL-2002; 2002US-0397755P.		
XX	PR	22-JUL-2002; 2002US-0397845P.		
XX	PR	09-SEP-2002; 2002US-0409450P.		
XX	PA	(EOSB-) EOS BIOTECHNOLOGY INC.		
XX	PI	Afar D, Aziz N, Ginsburg WM, Gish KC, Glynn R, Hevezi PA;		
XX	PI	Mack DH, Murray R, Watson SR, Wilson KE, Zlotnik A;		
XX	PI	WPI; 2003-468649/44.		
XX	DR	N-PSDB; ADN39301.		
XX	PT	Determining the presence or absence of a pathological cell in a patient,		
XX	PT	useful for diagnosing, prognosing or treating cancer, comprises detecting		
XX	PT	a nucleic acid in a biological sample.		
XX	PS	Claim 12; SEQ ID NO 620; 1385pp; English.		
XX	CC	The invention relates to nucleic acids and proteins (ADN39683-ADN40064)		
XX	CC	whose expression is upregulated or downregulated in specific cancers or		
XX	CC	other diseases such as angiogenic or fibrotic disorders, and to methods		
XX	CC	of determining the presence or absence of a pathological cell in a		
XX	CC	patient by detecting a nucleic acid at least 80% identical to those of		
XX	CC	the invention or by detecting a polypeptide of the invention. The		
XX	CC	invention also relates to expression vectors and host cells comprising a		
XX	CC	nucleic acid of the invention; antibodies which specifically bind a		
XX	CC	polypeptide of the invention; use of such antibodies for drug targeting;		
XX	CC	and methods of screening for modulators of activity or expression of the		
XX	CC	polypeptides and nucleic acids. The nucleic acids, polypeptides,		
XX	CC	antibodies and methods are useful for diagnosing, prognosing and treating		
XX	CC	cancer and other conditions such as psoriasis, ischaemia, heart disease,		
XX	CC	atherosclerosis, inflammatory diseases, autoimmune diseases, retinal		
XX	CC	neovascularisation syndromes, scarring and uterine fibroids. They may		
XX	CC	also be useful in wound healing and in contraception. The present		
XX	CC	sequence represents a polypeptide of the invention.		
XX	SQ	Sequence 357 AA;		
QY	1	MATTVPDGCRCNGLSKYYRLCDKAEAWGIVLETVATAGVWTSVAFMLTLPILVCKVQDSN	60	
Db	1	MATTVPDGCRCNGLSKYYRLCDKAEAWGIVLETVATAGVWTSVAFMLTLPILVCKVQDSN	60	
QY	61	RRKMLPTQFLFLLGVLGIFGLTFAPFIIGLDGSTGTRFRLFGILFSCILLAHAVSLT	120	
Db	61	RRKMLPTQFLFLLGVLGIFGLTFAPFIIGLDGSTGTRFRLFGILFSCILLAHAVSLT	120	
QY	121	KLVRGKPLSLVILGLAVGFSLVQDVIAEIVILTMRTNNVFSSELSAPRNEDFVLL	180	
Db	121	KLVRGKPLSLVILGLAVGFSLVQDVIAEIVILTMRTNNVFSSELSAPRNEDFVLL	180	
QY	181	LTYYVLFMLALTFLMSSFTFCGSFTGKRGHAIYLTMLLSIAIWAWITLLMLPDPDRW	240	
Db	181	LTYYVLFMLALTFLMSSFTFCGSFTGKRGHAIYLTMLLSIAIWAWITLLMLPDPDRW	240	
QY	241	DDTILSSALAANGWVFLLAYVSPFELWLTQKRNPDYVEDAFCKPQLVKKSYGVENRAY	300	
Db	241	DDTILSSALAANGWVFLLAYVSPFELWLTQKRNPDYVEDAFCKPQLVKKSYGVENRAY	300	
QY	301	SOEITQGFEEGTDTLYAPYSTHFOLQNPQPKFSPRAHAWPSPYKDYEVKKEGS	357	
Db	301	SOEITQGFEEGTDTLYAPYSTHFOLQNPQPKFSPRAHAWPSPYKDYEVKKEGS	357	
RESULT 9				
AD128460				
ID	AD128460	standard; protein; 357 AA.		
XX	AC	AD128460;		
XX	DT	22-APR-2004 (first entry)		
XX	DE	Human GPCR retinoic acid induced 3 (RAI-3) protein.		

XX Retinoic acid induced 3; RAI-3; human; G-protein coupled receptor; GPCR;  
KW antiinflammatory; immunosuppressive; cytotstatic; cardiant; antiallergic;  
KW broncholytic; gene therapy.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT Region 1..357  
FT /note= "Polypeptide specifically referred to in Claim 5 "  
FT Region 90..104  
FT /note= "Polypeptide specifically referred to in Claim 5 "  
FT Region 155..178  
FT /note= "Polypeptide specifically referred to in Claim 5 "  
FT Region 232..248  
FT /note= "Polypeptide specifically referred to in Claim 5 "  
FT Region 269..284  
FT /note= "Polypeptide specifically referred to in Claim 5 "  
FT Region 285..304  
FT /note= "Polypeptide specifically referred to in Claim 5 "  
FT Region 340..353  
FT /note= "Polypeptide specifically referred to in Claim 5 "  
XX WO2004001060-A2.  
PN 31-DEC-2003.  
XX 20-JUN-2003; 2003WO-US019255.  
XX 20-JUN-2002; 2002US-0390850P.  
PR 29-AUG-2002; 2002US-0407006P.  
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.  
XX Whitney GS, Opitack G, Garulacan L, Ramanathan CS, Mckinnon M;  
PI Bennett KL, Barber LE, Cacace A, Tauchihaashi Z;  
XX WPI: 2004-090973/09.  
DR N-PSDB; AD128459.  
XX New nucleic acid molecule encoding a human G-protein coupled receptor,  
PT RAI3, useful for preventing, treating or ameliorating chronic obstructive  
PT pulmonary disease (COPD), COPD-like disorder, or the underlying symptoms  
PT of COPD.  
XX Claim 5; SEQ ID NO 3; 301pp; English.  
XX The present sequence is the protein sequence of a human G-protein coupled  
CC receptor (Class C, Group 5), denoted retinoic acid induced 3 (RAI-3).  
CC Proteomics methods were used to isolate cigarette smoke-inducible  
CC tyrosine phosphorylated proteins from airway epithelial cells. RAI-3 was  
CC identified as being tyrosine phosphorylated and/or as being  
CC associated/complexed with tyrosine phosphorylated proteins only in cells  
CC that had been exposed to cigarette smoke. Since RAI-3 is primarily  
CC expressed in lung tissue, and since cigarette smoke is a major causative  
CC factor of chronic obstructive pulmonary disease (COPD), RAI-3 provides a  
CC novel cellular target for identifying modulators, e.g. agonists or  
CC antagonists, useful for the treatment and/or prevention of COPD and  
CC related disorders such as emphysema and chronic bronchitis. RAI-3  
CC modulators, e.g. agonists and antagonists, especially antisenese  
CC compounds, can be used to treat COPD and other disorders and diseases  
CC associated with regulation of NF-kB and/or its associated or interacting  
CC signalling molecules. Single nucleotide polymorphisms (SNPs) detected in  
CC the RAI-3 gene are useful for determining COPD association in  
CC individuals. 3 Of the SNPs result in alterations in the RAI-3 amino acid  
CC sequence. Of these, Ser118Gly and Thr182A are at positions which are not  
CC conserved between human and murine RAI-3 sequences, whereas Gln307Arg  
CC involves a conserved amino acid residue. RAI-3 nucleic acid molecules and  
CC polypeptides are useful for preventing, treating or ameliorating  
CC disorders related to aberrant GPCR signaling or cell cycle regulation,  
CC pulmonary disorders, inflammatory lung disorders, COPD, the underlying  
CC symptoms of COPD, COPD-related disorders or conditions, autoimmune  
CC disorders, disorders related to hyperimmune activity, inflammatory

CC conditions, disorders related to aberrant acute phase responses,  
CC hypercongenital conditions, birth defects, necrotic lesions, wounds,  
CC organ transplant rejection, renal diseases, ischaemia-reperfusion injury,  
CC heart disorders, disorders related to aberrant signal transduction,  
CC proliferation disorders, cancers, HIV infection, asthma, cystic fibrosis,  
CC pulmonary fibrosis, ulcerative colitis, cerebral infarct, myocardial  
CC infarct, diabetic nephropathy, allergic rhinitis, Crohn's disease,  
CC atherosclerosis, rheumatoid arthritis, inflammatory/autoimmune disorders,  
CC glioblastoma, pulmonary small cell undifferentiated carcinoma, carcinoma  
CC of the breast, colon, lung, ovary, pancreas, prostate, non-Hodgkin's  
CC lymphoma, disorders associated with aberrant cell adhesion, I-CAM  
CC function and/or regulation, E-selectin function and/or regulation, or  
CC aberrant NF-kB function and/or regulation (all claimed).  
XX Sequence 357 AA;  
SQ  
Query Match 100.0%; Score 1865; DB 8; Length 357;  
Best Local Similarity 100.0%; Pred. No. 2.7e-203;  
Matches 357; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MATTVPDGCGRNGLSKYYRLCDKAEAWGIVLETATAGVVTSAFMTLTPILVCKVQDSN 60  
DB 1 MATTVPDGCGRNGLSKYYRLCDKAEAWGIVLETATAGVVTSAFMTLTPILVCKVQDSN 60  
QY 61 RRKMLPTQFLFLGLVGLIFGLTFAPFTIGLDGSGTGRFFELFGLIFSCSLLAHAVSLT 120  
DB 61 RRKMLPTQFLFLGLVGLIFGLTFAPFTIGLDGSGTGRFFELFGLIFSCSLLAHAVSLT 120  
QY 121 KLVGRKPLSLVILGLAVGFSLVQDVIAIEYIVLTMTNNTNVNVSFELSAPRRNEDFVLL 180  
DB 121 KLVGRKPLSLVILGLAVGFSLVQDVIAIEYIVLTMTNNTNVNVSFELSAPRRNEDFVLL 180  
QY 181 LTYVFLMALTFMLSSFTFCGSGFTGKRGHAIHYLTMLLSIAIWAIVITLLMLPDDRWR 240  
DB 181 LTYVFLMALTFMLSSFTFCGSGFTGKRGHAIHYLTMLLSIAIWAIVITLLMLPDDRWR 240  
QY 241 DDTILSSALAANGWVFLLAYVSPFLLTKQNPMDYPVEDAFCKPQLVKKSQGVENRAY 300  
DB 241 DDTILSSALAANGWVFLLAYVSPFLLTKQNPMDYPVEDAFCKPQLVKKSQGVENRAY 300  
QY 301 SQEETITQGEETGDTLYAPYSTHFLQNPQPKQESIPRAHAWPSPYKDYEVKKEGS 357  
DB 301 SQEETITQGEETGDTLYAPYSTHFLQNPQPKQESIPRAHAWPSPYKDYEVKKEGS 357  
RESULT 10  
ADO29632  
ID ADO29632 standard; protein; 357 AA.  
XX  
AC ADO29632;  
XX  
DT 29-JUL-2004 (first entry)  
XX  
DE Human GPCR RAI3, SEQ ID NO:734.  
XX  
KW G protein-coupled receptor; GPCR; drug screening; diagnosis;  
KW transgenic mouse; neurological disorder; adrenal gland disorder;  
KW colon disorder; intestinal disorder; cardiovascular disorder;  
KW muscular disorder; blood disorder; immune disorder; bone disorder;  
KW joint disorder; metabolic disorder; nutritive disorder; cancer;  
KW kidney disorder; liver disorder; lung disorder; breast disorder;  
KW ovary disorder; uterus disorder; prostate disorder; testis disorder;  
KW skin disorder; stomach disorder; pancreas disorder; spleen disorder;  
KW thymus disorder; thyroid disorder; antiparkinsonian; antineptic;  
KW cytostatic; antiinflammatory; vasotropic; antiarrhythmic;  
KW CNS; central nervous system; respiratory; antidiarrhoeic; antidiabetic;  
KW virucide; hepatotropic; antibacterial; antianaemic; antiseborrhoeic;  
KW dermatological; antiulcer; antithyroid; antiallergic; anorectic;  
KW immunosuppressive; nephrotropic; gene therapy; GPCR modulator; human;  
KW receptor.  
XX Homo sapiens.  
XX



CC sequence represents human retinoic acid induced 3 (RAI3), which is used  
CC in the exemplification of the present invention.  
XX  
SQ Sequence 357 AA;  
  
Query Match 100.0%; Score 1865; DB 8; Length 357;  
Best Local Similarity 100.0%; Pred. No. 2.7e-203;  
Matches 357; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 MATTVPDGCGRGLKSKYYRLCDKAEAGIVLETVATAGVTVTSVAFMLTLPILVCKVQDSN 60  
DB |||||||  
1 MATTVPDGCGRGLKSKYYRLCDKAEAGIVLETVATAGVTVTSVAFMLTLPILVCKVQDSN 60  
QY 61 RRKMLPTQFLFLLGVLFGLTFAFIIGLDGSGTGRFFLFGILFSCFSCLLAHAVSLT 120  
DB |||||||  
61 RRKMLPTQFLFLLGVLFGLTFAFIIGLDGSGTGRFFLFGILFSCFSCLLAHAVSLT 120  
QY 121 KLVRGRKPLSLVILGLAVGFSLVQDVIAIEYIVLTMTNRTNVVSELSAPRRNEDFVLL 180  
DB |||||||  
121 KLVRGRKPLSLVILGLAVGFSLVQDVIAIEYIVLTMTNRTNVVSELSAPRRNEDFVLL 180  
QY 181 LTYVFLMALTFMLMSSFTFCGSGFTGKRGHAIYLTMLLSIAIWAWITLLMLPDPDRRW 240  
DB |||||||  
181 LTYVFLMALTFMLMSSFTFCGSGFTGKRGHAIYLTMLLSIAIWAWITLLMLPDPDRRW 240  
QY 241 DDTILSSALAANGWVFLAYVSPFELTQKRNPMQPVVEDAFCKPQLVKKS YGVENRAY 300  
DB |||||||  
241 DDTILSSALAANGWVFLAYVSPFELTQKRNPMQPVVEDAFCKPQLVKKS YGVENRAY 300  
QY 301 SQEEITQGFEEGTDTLYAPYSTHFQLOQPPOKEFSIPRAHAWPSYKDYEVKKEGS 357  
DB |||||||  
301 SQEEITQGFEEGTDTLYAPYSTHFQLOQPPOKEFSIPRAHAWPSYKDYEVKKEGS 357  
  
RESULT 12  
ABM81354  
ID ABM81354 standard; protein; 357 AA.  
XX  
AC ABM81354;  
XX  
DT 18-NOV-2004 (first entry)  
XX  
DE Tumour-associated antigenic target (TAT) polypeptide PRO37234, SEQ:3499.  
XX  
KW Tumour-associated antigenic target; TAT; human; overexpression; cancer;  
KW tumour; diagnosis; cell proliferative disorder; breast cancer;  
KW colorectal cancer; lung cancer; ovarian cancer; liver cancer;  
KW central nervous system cancer; bladder cancer; pancreatic cancer;  
KW cervical cancer; melanoma; leukaemia; hybridisation probe;  
KW chromosome identification; chromosome mapping; gene mapping;  
KW gene therapy; cytostatic.  
XX  
OS Homo sapiens.  
XX  
PN WO2004030615-A2.  
XX  
PD 15-APR-2004.  
XX  
PF 29-SEP-2003; 2003WO-US028547.  
XX  
PR 02-OCT-2002; 2002US-0414971P.  
XX  
PA (GETH ) GENENTECH INC.  
XX  
PI Wu TD, Zhang Z, Zhou Y;  
XX  
PI WPI; 2004-347921/32.  
XX  
DR N-PSDB; ACN39388.  
XX  
PT New tumor-associated antigenic target polypeptides and nucleic acids,  
PT useful in preparing a medicament for treating or detecting a  
PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or  
PT prostate cancer or tumor.

XX  
PS  
XX  
CC Claim 12; SEQ ID NO 3499; 7273pp; English.  
CC The invention relates to human tumour-associated antigenic target (TAT)  
CC polypeptides, and their related nucleic acids. The TAT polypeptides are  
CC overexpressed in cancer tissues compared to normal tissues, and may thus  
CC serve as effective targets for the diagnosis and treatment of cancer in  
CC mammals. The invention also relates to nucleic acid and polypeptide  
CC sequences at least 80% identical to the TAT nucleic acids and  
CC polypeptides; expression vectors and host cells comprising a TAT nucleic  
CC acid; an antibody specific for a TAT polypeptide; a peptide or organic  
CC molecule which binds to a TAT polypeptide; fusion proteins comprising a  
CC TAT polypeptide; and methods and compositions for the treatment or  
CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,  
CC antibodies, antagonists, binding molecules and compositions are useful  
CC for diagnosing or treating a cell proliferative disorder associated with  
CC increased TAT expression, particularly cancers such as breast cancer,  
CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder  
CC cancer, pancreatic cancer, cervical cancer, cancers of the central  
CC nervous system, melanoma and leukaemia. TAT nucleic acids may further be  
CC used as hybridisation probes, in chromosome and gene mapping, in  
CC chromosome identification and in gene therapy. The present sequence  
CC represents a TAT polypeptide of the invention  
XX  
SQ Sequence 357 AA;  
  
Query Match 100.0%; Score 1865; DB 8; Length 357;  
Best Local Similarity 100.0%; Pred. No. 2.7e-203;  
Matches 357; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 MATTVPDGCGRGLKSKYYRLCDKAEAGIVLETVATAGVTVTSVAFMLTLPILVCKVQDSN 60  
DB |||||||  
1 MATTVPDGCGRGLKSKYYRLCDKAEAGIVLETVATAGVTVTSVAFMLTLPILVCKVQDSN 60  
QY 61 RRKMLPTQFLFLLGVLFGLTFAFIIGLDGSGTGRFFLFGILFSCFSCLLAHAVSLT 120  
DB |||||||  
61 RRKMLPTQFLFLLGVLFGLTFAFIIGLDGSGTGRFFLFGILFSCFSCLLAHAVSLT 120  
QY 121 KLVRGRKPLSLVILGLAVGFSLVQDVIAIEYIVLTMTNRTNVVSELSAPRRNEDFVLL 180  
DB |||||||  
121 KLVRGRKPLSLVILGLAVGFSLVQDVIAIEYIVLTMTNRTNVVSELSAPRRNEDFVLL 180  
QY 181 LTYVFLMALTFMLMSSFTFCGSGFTGKRGHAIYLTMLLSIAIWAWITLLMLPDPDRRW 240  
DB |||||||  
181 LTYVFLMALTFMLMSSFTFCGSGFTGKRGHAIYLTMLLSIAIWAWITLLMLPDPDRRW 240  
QY 241 DDTILSSALAANGWVFLAYVSPFELTQKRNPMQPVVEDAFCKPQLVKKS YGVENRAY 300  
DB |||||||  
241 DDTILSSALAANGWVFLAYVSPFELTQKRNPMQPVVEDAFCKPQLVKKS YGVENRAY 300  
QY 301 SQEEITQGFEEGTDTLYAPYSTHFQLOQPPOKEFSIPRAHAWPSYKDYEVKKEGS 357  
DB |||||||  
301 SQEEITQGFEEGTDTLYAPYSTHFQLOQPPOKEFSIPRAHAWPSYKDYEVKKEGS 357  
  
RESULT 13  
ADU06613  
ID ADU06613 standard; protein; 357 AA.  
XX  
AC ADU06613;  
XX  
DT 27-JAN-2005 (first entry)  
XX  
DE Novel bronchial cancer-associated human protein SeqID839.  
XX  
KW bronchial cancer; cytostatic; tumour-associated protein;  
KW cancer detection; metastasis; tumour; human.  
XX  
OS Homo sapiens.  
XX  
PN DE10316701-Al.  
XX  
PD 04-NOV-2004.





DB 1 MATTVPGCRNGLSKYYRLCDKAEAGVILETVATAGVVTSAFVMTLPILVCKVQDSN 60  
QY 61 RRKMLPTQFLFLLGVIGFGLTFAFIIGLDGSGTPTFRFELGILFSCFSCLLAHAVSLT 120  
DB 61 RRKMLPTQFLFLLGVIGFGLTFAFIIGLDGSGTPTFRFELGILFSCFSCLLAHAVSLT 120  
QY 121 KLVGRKPLSLVILGLAVGFSLVQDVIAEIVLTMNRTNNTNVSSELSAPRNEDFVLL 180  
DB 121 KLVGRKPLSLVILGLAVGFSLVQDVIAEIVLTMNRTNNTNVSSELSAPRNEDFVLL 180  
QY 181 LTYVFLMALTFELMSSFTFCGSGTGWKRHGAHIYLTMLLSIAIWAVITLLMLPDPDRW 240  
DB 181 LTYVFLMALTFELMSSFTFCGSGTGWKRHGAHIYLTMLLSIAIWAVITLLMLPDPDRW 240  
QY 241 DDTILSSALAANGWVFLAYVSPFWLLTKQRPMDYPVEDAFCKPQLVKKSYGVENRAY 300  
DB 241 DDTILSSALAANGWVFLAYVSPFWLLTKQRPMDYPVEDAFCKPQLVKKSYGVENRAY 300  
QY 301 SOEITQGFEEETGDTLYAPYSTHFOLQNPPOKEFSIPRAHAWPSYKDYEVKKEGS 357  
DB 301 SOEITQGFEEETGDTLYAPYSTHFOLQNPPOKEFSIPRAHAWPSYKDYEVKKEGS 357

RESULT 15  
AEA00190  
ID AEA00190 standard; protein; 357 AA.  
XX AEA00190;  
AC  
DT 28-JUL-2005 (first entry)  
TX  
DE Human TAT143 protein SEQ ID NO:142.  
XX  
KW tumor-associated antigen; cytostatic; breast tumor; endocrine disease;  
KW gynecology and obstetrics; neoplasm; colon tumor;  
KW gastrointestinal disease; rectal tumor; endometroid carcinoma;  
KW genitourinary disease; renal tumor; lung tumor; respiratory disease;  
KW ovary tumor; skin tumor; liver tumor.  
XX  
OS Homo sapiens.  
XX  
PN US2005106644-A1.  
XX  
PD 19-MAY-2005.  
XX  
PF 08-SEP-2004; 2004US-00936626.  
XX  
PR 20-JUN-2001; 2001US-0299500P.  
PR 29-JUN-2001; 2001US-0301880P.  
PR 18-SEP-2001; 2001US-0323268P.  
PR 19-JUN-2002; 2002US-0017748P.  
PR 26-MAR-2004; 2004US-0557116P.  
PR 04-AUG-2004; 2004US-0598899P.  
XX  
PA (GETH ) GENENTECH INC.  
XX  
XX Cairns B, Chen R, Frantz G, Hillan KJ, Koeppen H, Phillips HS;  
PI Polakis P, Spencer SD, Smith V, Williams PM, Wu TD, Zhang Z;  
PI Sliwkowski M;  
XX  
DR WPI; 2005-384304/39.  
DR N-PSDB; AEA00112.  
XX  
XX Novel isolated antibody capable of binding to tumor-associated antigenic  
PT target polypeptide, useful for treating cell proliferative disorder e.g.  
PT cancer.  
XX  
XX Claim 1; SEQ ID NO 142; 337pp; English.  
XX  
XX The invention relates to a novel isolated antibody binding to a  
CC polypeptide having at least 80% sequence identity to a polypeptide having  
CC any one of 76 fully defined 182-910 amino acid tumor-associated antigenic

CC target polypeptide (TAT) sequences (AEA00127-AEA00202) given in the  
CC specification, a polypeptide having any one of (AEA00127-AEA00202),  
CC lacking its associated signal peptide, or an extracellular domain of a  
CC polypeptide having any one of (AEA00127-AEA00202). The polypeptide is  
CC encoded by the nucleotide sequence having any one of (AEA00049-AEA00126).  
CC An antibody of the invention has cytostatic activity. The antibody is  
CC useful for inhibiting growth of a cell expressing TAT188, which involves  
CC contacting the cell with the antibody. The cell is a cancer cell chosen  
CC from breast, colon, rectum, endometrium, kidney, lung, ovary, skin and  
CC liver cell. The cancer cell is a mammalian cell, preferably a human cell.  
CC The antibody is also useful for detecting the level of TAT188  
CC polypeptide expressed in a test cell relative to a control cell, and for  
CC detecting the level of TAT188 polypeptide or a polypeptide having at  
CC least 80 % sequence identity to the TAT188 polypeptide sequence in a test  
CC cell relative to a control cell. The antibody is useful for treating a  
CC preventing a cell proliferative disorder associated with increased  
CC expression or activity of a polypeptide having at least 80 % identity to  
CC a TAT188 polypeptide sequence. The cell proliferative disorder is cancer.  
CC The method of the invention is useful for inhibiting the growth of a  
CC cancer cell. The present sequence represents a polypeptide of the  
CC invention.  
XX  
SQ Sequence 357 AA;  
Query Match 100.0%; Score 1865; DB 9; Length 357;  
Best Local Similarity 100.0%; Pred. No. 2.7e-203;  
Matches 357; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MATTVPGCRNGLSKYYRLCDKAEAGVILETVATAGVVTSAFVMTLPILVCKVQDSN 60  
DB 1 MATTVPGCRNGLSKYYRLCDKAEAGVILETVATAGVVTSAFVMTLPILVCKVQDSN 60  
QY 61 RRKMLPTQFLFLLGVIGFGLTFAFIIGLDGSGTPTFRFELGILFSCFSCLLAHAVSLT 120  
DB 61 RRKMLPTQFLFLLGVIGFGLTFAFIIGLDGSGTPTFRFELGILFSCFSCLLAHAVSLT 120  
QY 121 KLVGRKPLSLVILGLAVGFSLVQDVIAEIVLTMNRTNNTNVSSELSAPRNEDFVLL 180  
DB 121 KLVGRKPLSLVILGLAVGFSLVQDVIAEIVLTMNRTNNTNVSSELSAPRNEDFVLL 180  
QY 181 LTYVFLMALTFELMSSFTFCGSGTGWKRHGAHIYLTMLLSIAIWAVITLLMLPDPDRW 240  
DB 181 LTYVFLMALTFELMSSFTFCGSGTGWKRHGAHIYLTMLLSIAIWAVITLLMLPDPDRW 240  
QY 241 DDTILSSALAANGWVFLAYVSPFWLLTKQRPMDYPVEDAFCKPQLVKKSYGVENRAY 300  
DB 241 DDTILSSALAANGWVFLAYVSPFWLLTKQRPMDYPVEDAFCKPQLVKKSYGVENRAY 300  
QY 301 SOEITQGFEEETGDTLYAPYSTHFOLQNPPOKEFSIPRAHAWPSYKDYEVKKEGS 357  
DB 301 SOEITQGFEEETGDTLYAPYSTHFOLQNPPOKEFSIPRAHAWPSYKDYEVKKEGS 357  
Search completed: June 8, 2006, 10:27:24  
Job time : 204 secs

**THIS PAGE IS BLANK**



QY 301 SQEETQGFETGDTLYAPYSTHFOQNPPQKFSIPRAHAWPSPYKDYEVKKEGS 357  
|||||  
Db 301 SQEETQGFETGDTLYAPYSTHFOQNPPQKFSIPRAHAWPSPYKDYEVKKEGS 357  
|||||  
RESULT 2  
US-09-949-016-9484  
; Sequence 9484, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; CURRENT FILING DATE: 2000-04-14  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 9484  
; LENGTH: 390  
; TYPE: PRT  
; ORGANISM: Human  
US-09-949-016-9484

Query Match 100.0%; Score 1865; DB 2; Length 390;  
Best Local Similarity 100.0%; Pred. No. 3.7e-194;  
Matches 357; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATTPDGCGRGLKSKYRLCDKAEAWGIVLETATAGVWTSVAFMLTLPILVCKVQDSN 60  
|||||  
Db 34 MATTPDGCGRGLKSKYRLCDKAEAWGIVLETATAGVWTSVAFMLTLPILVCKVQDSN 93  
|||||  
QY 61 RRKMLPTQFLFLGLVGLFGLTFAFIIGLDGSTGTRFPLFGILFISICFSCLLAHAVSLT 120  
|||||  
Db 94 RRKMLPTQFLFLGLVGLFGLTFAFIIGLDGSTGTRFPLFGILFISICFSCLLAHAVSLT 153  
|||||  
QY 121 KLVRGRKPLSLVILGLAVGFLSDVDIAIEYIVLTMTNTNVNVSSELSAPRNEDFVLL 180  
|||||  
Db 154 KLVRGRKPLSLVILGLAVGFLSDVDIAIEYIVLTMTNTNVNVSSELSAPRNEDFVLL 213  
|||||  
QY 181 LTYVLFMLATFLMSSTFCGSGFTGKRGHAIYLTMLLSIAIWAIVITLLMLPDPDRRW 240  
|||||  
Db 214 LTYVLFMLATFLMSSTFCGSGFTGKRGHAIYLTMLLSIAIWAIVITLLMLPDPDRRW 273  
|||||  
QY 241 DDTILSSALAANGWFLLAYVSPFLLTKORNPMDYPVEDAFCKPQLVKKSYGVENRAY 300  
|||||  
Db 274 DDTILSSALAANGWFLLAYVSPFLLTKORNPMDYPVEDAFCKPQLVKKSYGVENRAY 333  
|||||  
QY 301 SQEETQGFETGDTLYAPYSTHFOQNPPQKFSIPRAHAWPSPYKDYEVKKEGS 357  
|||||  
Db 334 SQEETQGFETGDTLYAPYSTHFOQNPPQKFSIPRAHAWPSPYKDYEVKKEGS 390  
|||||

RESULT 3  
US-09-188-930-326  
; Sequence 326, Application US/09188930A  
; Patent No. 6150502  
; GENERAL INFORMATION:  
; APPLICANT: Watson, James D.  
; APPLICANT: Strachan, Lorna  
; APPLICANT: Sleeman, Matthew  
; APPLICANT: Onrust, Rene  
; APPLICANT: Murison, James Greg  
; TITLE OF INVENTION: Compositions Isolated From Skin Cells  
; FILE REFERENCE: 11000.1011c1  
; CURRENT APPLICATION NUMBER: US/09/188,930A

; CURRENT FILING DATE: 1998-11-09  
; NUMBER OF SEQ ID NOS: 348  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 326  
; LENGTH: 347  
; TYPE: PRT  
; ORGANISM: Human  
US-09-188-930-326

Query Match 95.5%; Score 1782; DB 2; Length 347;  
Best Local Similarity 99.1%; Pred. No. 3.3e-185;  
Matches 341; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 14 KSKYYRLCDKAEAWGIVLETATAGVWTSVAFMLTLPILVCKVQDSNRRKMLPTQFLFLL 73  
: : : : :  
Db 4 RPRYYRLCDKAEAWGIVLETATAGVWTSVAFMLTLPILVCKVQDSNRRKMLPTQFLFLL 63  
: : : : :  
QY 74 GVLGIFGLTFAFIIGLDGSTGTRFPLFGILFISICFSCLLAHAVSLTKLVRGRKPLSLV 133  
|||||  
Db 64 GVLGIFGLTFAFIIGLDGSTGTRFPLFGILFISICFSCLLAHAVSLTKLVRGRKPLSLV 123  
|||||

QY 134 ILGLAVGFLSDVDIAIEYIVLTMTNTNVNVSSELSAPRNEDFVLLTYVLFMALTF 193  
|||||  
Db 124 ILGLAVGFLSDVDIAIEYIVLTMTNTNVNVSSELSAPRNEDFVLLTYVLFMALTF 183  
|||||

QY 194 MSSFTFCGSGFTGKRGHAIYLTMLLSIAIWAIVITLLMLPDPDRWDTTILSSALAANG 253  
|||||  
Db 184 MSSFTFCGSGFTGKRGHAIYLTMLLSIAIWAIVITLLMLPDPDRWDTTILSSALAANG 243  
|||||

QY 254 WFLLAYVSPFLLTKORNPMDYPVEDAFCKPQLVKKSYGVENRAYSQEETQGFETG 313  
|||||  
Db 244 WFLLAYVSPFLLTKORNPMDYPVEDAFCKPQLVKKSYGVENRAYSQEETQGFETG 303  
|||||

QY 314 DTLYAPYSTHFOQNPPQKFSIPRAHAWPSPYKDYEVKKEGS 357  
|||||  
Db 304 DTLYAPYSTHFOQNPPQKFSIPRAHAWPSPYKDYEVKKEGS 347  
|||||

RESULT 4  
US-09-312-283C-326  
; Sequence 326, Application US/09312283C  
; Patent No. 6573095  
; GENERAL INFORMATION:  
; APPLICANT: Watson, James D.  
; APPLICANT: Strachan, Lorna  
; APPLICANT: Sleeman, Matthew  
; APPLICANT: Onrust, Rene  
; APPLICANT: Murison, James G.  
; APPLICANT: Kumble, Krishanand D.  
; TITLE OF INVENTION: Compositions Isolated from Skin Cells  
; FILE REFERENCE: 11000.1011c2  
; CURRENT APPLICATION NUMBER: US/09/312,283C  
; CURRENT FILING DATE: 1999-05-14  
; NUMBER OF SEQ ID NOS: 425  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 326  
; LENGTH: 347  
; TYPE: PRT  
; ORGANISM: Mouse  
US-09-312-283C-326

Query Match 95.5%; Score 1782; DB 2; Length 347;  
Best Local Similarity 99.1%; Pred. No. 3.3e-185;  
Matches 341; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 14 KSKYYRLCDKAEAWGIVLETATAGVWTSVAFMLTLPILVCKVQDSNRRKMLPTQFLFLL 73  
: : : : :  
Db 4 RPRYYRLCDKAEAWGIVLETATAGVWTSVAFMLTLPILVCKVQDSNRRKMLPTQFLFLL 63  
: : : : :  
QY 74 GVLGIFGLTFAFIIGLDGSTGTRFPLFGILFISICFSCLLAHAVSLTKLVRGRKPLSLV 133  
|||||  
Db 64 GVLGIFGLTFAFIIGLDGSTGTRFPLFGILFISICFSCLLAHAVSLTKLVRGRKPLSLV 123  
|||||

QY 134 ILGLAVGFSIVODVIAIEYIVLTMTNNTNNVSELSAPRNEDFVLLLTVVLFMALTEL 193  
DB 124 ILGLAVGFSIVODVIAIEYIVLTMTNNTNNVSELSAPRNEDFVLLLTVVLFMALTEL 183  
QY 194 MSSFTFCGSGFTGKRGHAIYLTMLLSIAIWAVITLLMLPDRDWDITLSSALAANG 253  
DB 184 MSSFTFCGSGFTGKRGHAIYLTMLLSIAIWAVITLLMLPDRDWDITLSSALAANG 243  
QY 254 WFLLAYVSPFWLLTKQRNPDYFVEDAFCKPQLVKYSYGVENRAYSOEETQGPETG 313  
DB 244 WFLLAYVSPFWLLTKQRNPDYFVEDAFCKPQLVKYSYGVENRAYSOEETQGPETG 303  
QY 314 DTLVAPYSTHFQLOQNPQKESIPRAHAWPSYKDYEVKKEGS 357  
DB 304 DTLVAPYSTHFQLOQNPQKESIPRAHAWPSYKDYEVKKEGS 347

## RESULT 5

US-09-188-930-123  
; Sequence 123, Application US/09188930A  
; Patent No. 6150502  
; GENERAL INFORMATION:  
; APPLICANT: Watson, James D.  
; APPLICANT: Strachan, Lorna  
; APPLICANT: Sleeman, Matthew  
; APPLICANT: Onrust, Rene  
; APPLICANT: Murison, James G.  
; TITLE OF INVENTION: Compositions Isolated From Skin Cells  
; FILE REFERENCE: 11000.1011c1  
; CURRENT APPLICATION NUMBER: US/09/188,930A  
; CURRENT FILING DATE: 1998-11-09  
; NUMBER OF SEQ ID NOS: 348  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 123  
; LENGTH: 68  
; TYPE: PRT  
; ORGANISM: Human  
US-09-188-930-123

Query Match 18.8%; Score 350; DB 2; Length 68;  
Best Local Similarity 100.0%; Pred. No. 2e-30;  
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 MTLPLVCKVQDSNRRKMLPTQFLFLGLVGLFGLTFAFIICLDGSGTGTPTFFFLGILF 105  
DB 1 MTLPLVCKVQDSNRRKMLPTQFLFLGLVGLFGLTFAFIICLDGSGTGTPTFFFLGILF 60

QY 106 SICFSCLL 113  
DB 61 SICFSCLL 68

## RESULT 6

US-09-312-283C-123  
; Sequence 123, Application US/09312283C  
; Patent No. 6573095  
; GENERAL INFORMATION:  
; APPLICANT: Watson, James D.  
; APPLICANT: Strachan, Lorna  
; APPLICANT: Sleeman, Matthew  
; APPLICANT: Onrust, Rene  
; APPLICANT: Murison, James G.  
; APPLICANT: Kumble, Krishanand D.  
; TITLE OF INVENTION: Compositions Isolated from Skin Cells  
; FILE REFERENCE: 11000.1011c2  
; CURRENT APPLICATION NUMBER: US/09/312,283C  
; CURRENT FILING DATE: 1999-05-14  
; NUMBER OF SEQ ID NOS: 425  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 123

; LENGTH: 68  
; TYPE: PRT  
; ORGANISM: Mouse  
US-09-312-283C-123

Query Match 18.8%; Score 350; DB 2; Length 68;  
Best Local Similarity 100.0%; Pred. No. 2e-30;  
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 MTLPLVCKVQDSNRRKMLPTQFLFLGLVGLFGLTFAFIICLDGSGTGTPTFFFLGILF 105  
DB 1 MTLPLVCKVQDSNRRKMLPTQFLFLGLVGLFGLTFAFIICLDGSGTGTPTFFFLGILF 60

QY 106 SICFSCLL 113  
DB 61 SICFSCLL 68

## RESULT 7

US-09-964-956-85  
; Sequence 85, Application US/09964956  
; Patent No. 6875570  
; GENERAL INFORMATION:  
; APPLICANT: Gerlach, Valerie L  
; APPLICANT: MacDougall, John R  
; APPLICANT: Smithson, Glennda  
; APPLICANT: Millet, Isabelle  
; APPLICANT: Stone, David  
; APPLICANT: Gunther, Erik  
; APPLICANT: Ellerman, Karen  
; APPLICANT: Grosse, William M  
; APPLICANT: Alsobrook II, John P  
; APPLICANT: Lepley, Denise M  
; APPLICANT: Burgess, Catherine E  
; APPLICANT: Padigaru, Muralidhara  
; APPLICANT: Kekuda, Ramesh  
; APPLICANT: Spytek, Kimberly A  
; APPLICANT: Leach, Martin D  
; APPLICANT: Shimkets, Richard A  
; TITLE OF INVENTION: No. 6875570el Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-124  
; CURRENT APPLICATION NUMBER: US/09/964,956  
; CURRENT FILING DATE: 2001-09-26  
; PRIOR APPLICATION NUMBER: 60/235,631  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: 60/235,633  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: 60/235,808  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: 60/236,064  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: 60/236,065  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: 60/236,066  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: 60/236,135  
; PRIOR FILING DATE: 2000-09-28  
; PRIOR APPLICATION NUMBER: 60/237,434  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/238,321  
; PRIOR FILING DATE: 2000-10-05  
; PRIOR APPLICATION NUMBER: 60/238,399  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: 60/238,396  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: 60/276,667  
; PRIOR FILING DATE: 2001-03-16  
; PRIOR APPLICATION NUMBER: 60/294,823  
; PRIOR FILING DATE: 2001-05-31  
; PRIOR APPLICATION NUMBER: 60/304,868  
; PRIOR FILING DATE: 2001-07-12  
; NUMBER OF SEQ ID NOS: 127  
; SOFTWARE: PatentIn Ver. 2.1

US-09-964-956-85

```

; SEQ ID NO 85
; LENGTH: 256
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: 7
; OTHER INFORMATION: transmembrane receptor (metabotropic B family)
; OTHER INFORMATION: Consensus Sequence
US-09-964-956-85

Query Match      8.4%; Score 157.5; DB 2; Length 256;
Best Local Similarity 27.5%; Pred. No. 1.2e-08;
Matches 65; Conservative 43; Mismatches 75; Indels 53; Gaps 11;

QY 28 GIVLETVATAGVTVTVAFMLTLPILVCKVQD-----SNRRKMLPTQFLFLGLVGIF 79
Db 2 GIVLVALVGLIV-----LTLFVLVVFVRQDTPVIVKASNRE-----LSVLLILGLILCY 51
QY 80 GLTFFAIIGLDGSTGPT-----RFFLFGILFSCISCLLAHAVALSLTKLVGRGKPLS----- 130
Db 52 LCSFLPI-----GKPSETSCILARILFGLGFLCYLSALLAKTNVLRIRFRAKKPGSGKPKF 107
QY 131 -----LLVILGLAVGFSLVQDVIAIEYIVLTMTNNVNFSE-----LSAPRRN-BDFV 178
Db 108 ISPWAQVLIVLIV-----LIQIICVIMLVVEPPRPTDIYSEKEKILCEKNGKSMVAFV 163
QY 179 LLLTYVLFMAITFLMSSFT--FCGFTGKRGHGAHIYLTLLSLAIWVAWITLLM 232
Db 164 VILGYDGLLAVLCTFLAFLTRNLNPFNEAK-----FIGFSMLTFCIVWVAFIPIYL 215

RESULT 8
US-09-964-956-53
; Sequence 53, Application US/09964956
; Patent No. 6875570
; GENERAL INFORMATION:
; APPLICANT: Gerlach, Valerie L
; APPLICANT: MacDougall, John R
; APPLICANT: Smithson, Glenna
; APPLICANT: Millet, Isabelle
; APPLICANT: Stone, David
; APPLICANT: Gunther, Erik
; APPLICANT: Ellerman, Karen
; APPLICANT: Grosse, William M
; APPLICANT: Alsobrook II, John P
; APPLICANT: Lepley, Denise M
; APPLICANT: Burgess, Catherine E
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Leach, Martin D
; APPLICANT: Shinkets, Richard A
; TITLE OF INVENTION: No. 6875570el Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-124
; CURRENT APPLICATION NUMBER: US/09/964,956
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/235,631
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/235,633
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/235,808
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/236,064
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/236,065
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/236,066
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/236,135
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: 60/237,434
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/238,321

; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/238,399
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/238,396
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/276,667
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/294,823
; PRIOR FILING DATE: 2001-05-31
; PRIOR APPLICATION NUMBER: 60/304,868
; PRIOR FILING DATE: 2001-07-12
; NUMBER OF SEQ ID NOS: 127
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 53
; LENGTH: 879
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-964-956-53

Query Match      7.9%; Score 146.5; DB 2; Length 879;
Best Local Similarity 21.7%; Pred. No. 1.1e-06;
Matches 73; Conservative 57; Mismatches 148; Indels 59; Gaps 11;

QY 8 GCRNLGSKYYRLCDKAEAWGIVLETVATAGVTVTVAFMLTLPILVCKVQDSNR----- 61
Db 561 GCYN-LPEDYIRWED---AWAIGPVTTIACLG-----FMCTCIVITVFIKHNNTPLVKAS 610
QY 62 -RKMLPTQFLFLGLVGIFGLTFAFIIGLDGSTGPTFRFFLFGILFSCISCLLAHAVALSLT 120
Db 611 GRELC---YILLFGVLSYCMTFFFIAPKSPVICALRRRLGLGTSFAICYSALLTKNKCIA 667
QY 121 KLVGRGK-----PLSLLVILGLAVGFSLVQDVIAIEYIVLTMTNNVNFSE--- 167
Db 668 RIFDGVKNQAORPKFISPSQVFI---CLGLILVIVMWSWLLILETPGTRRYTLPEKRE 724
QY 168 ---LSAPRRNEDFVLLITY---VLFMLALTFLMSSFTFCGFTGKRGHGAHIYLTLLSLIA 222
Db 725 TVILKCNVKSMSLISLTIVDVVLVILCTVYAFKTRKCPENFNEAKFTGFTMYTTCI---- 780
QY 223 INVAWITLLMLPDPDRRWDITLSSALAANGWVFLAYVSPFEWLLTKQRNPMDPVEDA 282
Db 781 IWLAFPIFYVTSSDYRVQTTMCISVLSGFFVVLGCLFAPKVHIVLFQ----- 829
QY 283 FCKPOLVKKSYGVENRAYSQBEITQGFETGDTLIYAP 319
Db 830 ---PQKNVVTTHRLHLNRPFSVSGTATTYSQSSASTYVP 863

RESULT 9
US-09-964-956-54
; Sequence 54, Application US/09964956
; Patent No. 6875570
; GENERAL INFORMATION:
; APPLICANT: Gerlach, Valerie L
; APPLICANT: MacDougall, John R
; APPLICANT: Smithson, Glenna
; APPLICANT: Millet, Isabelle
; APPLICANT: Stone, David
; APPLICANT: Gunther, Erik
; APPLICANT: Ellerman, Karen
; APPLICANT: Grosse, William M
; APPLICANT: Alsobrook II, John P
; APPLICANT: Lepley, Denise M
; APPLICANT: Burgess, Catherine E
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Leach, Martin D
; APPLICANT: Shinkets, Richard A
; TITLE OF INVENTION: No. 6875570el Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-124
; CURRENT APPLICATION NUMBER: US/09/964,956
; CURRENT FILING DATE: 2001-09-26

```

;; PRIOR APPLICATION NUMBER: 60/235,631  
;; PRIOR FILING DATE: 2000-09-27  
;; PRIOR APPLICATION NUMBER: 60/235,633  
;; PRIOR FILING DATE: 2000-09-27  
;; PRIOR APPLICATION NUMBER: 60/235,808  
;; PRIOR FILING DATE: 2000-09-27  
;; PRIOR APPLICATION NUMBER: 60/236,064  
;; PRIOR FILING DATE: 2000-09-27  
;; PRIOR APPLICATION NUMBER: 60/236,065  
;; PRIOR FILING DATE: 2000-09-27  
;; PRIOR APPLICATION NUMBER: 60/236,066  
;; PRIOR FILING DATE: 2000-09-27  
;; PRIOR APPLICATION NUMBER: 60/236,135  
;; PRIOR FILING DATE: 2000-09-28  
;; PRIOR APPLICATION NUMBER: 60/237,434  
;; PRIOR FILING DATE: 2000-10-03  
;; PRIOR APPLICATION NUMBER: 60/238,321  
;; PRIOR FILING DATE: 2000-10-05  
;; PRIOR APPLICATION NUMBER: 60/238,399  
;; PRIOR FILING DATE: 2000-10-06  
;; PRIOR APPLICATION NUMBER: 60/238,396  
;; PRIOR FILING DATE: 2000-10-06  
;; PRIOR APPLICATION NUMBER: 60/276,667  
;; PRIOR FILING DATE: 2001-03-16  
;; PRIOR APPLICATION NUMBER: 60/294,823  
;; PRIOR FILING DATE: 2001-05-31  
;; PRIOR APPLICATION NUMBER: 60/304,868  
;; PRIOR FILING DATE: 2001-07-12  
;; NUMBER OF SEQ ID NOS: 127  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 54  
;; LENGTH: 879  
;; TYPE: PRT  
;; ORGANISM: Mus musculus  
US-09-964-956-54

Query Match 7.9%; Score 146.5; DB 2; Length 879;  
Best Local Similarity 21.7%; Pred. No. 1.1e-06;  
Matches 73; Conservative 57; Mismatches 148; Indels 59; Gaps 11;  
  
QY 8 GCRGLSKYRILCDKAEANGVLETATAGVVTSAFMTLTPILVCKVQDSNR----- 61  
DB 561 GCYN-LPEDYIRWED---AWAIGPVTIACLG-----FMCTCIVITVFIKHNTPLVKAS 610  
  
QY 62 -RKMLPTQFLFLGLVIGLTFAPFIIGLDGSTGPTFRFLFGILFSCIFSCLLAHAVSLT 120  
DB 611 GRELC---YILLFGVSLSYCMTEFFFAKPSVICALRLRLGLGTSFAICYSALLTKTNCA 667  
  
QY 121 KLVGRGK-----PLSLVILGLAVGFSLVQDVIAIEYIVLTMTNRTNVVFSE--- 167  
DB 668 RIFDGVNGAQRPKFISPSQVFI---CLGLILVQIVMVSVMLILETPGTRRYTLPEKRE 724  
  
QY 168 ---LSAPRNEDFVLLTY--VLFMLALTFLMSSFTFCGSGFTGKRGHGAHIYLTMLLSIA 222  
DB 725 TVILKCNVKDSSMLISLTYDVVLVILCTVYAFKTRKCPENFNEAKFIGFTMYTTCI--- 780  
  
QY 223 IIVAVITLMLPDRRDDDTLSSALANGWVFLAYSPFELWLTQKRNPMQNDYVEDA 282  
DB 781 IWLAFPLFIYVTSRDRVOTTTMCISVLSGFFVLGCLFAPKVHILFQ----- 829  
  
QY 283 FCKPOLVKSGYVENRAYSQEEITQGFETGDTLYAP 319  
DB 830 ---PQKNVVTHRLHNRFSVSGTATYSQSSASTYVP 863

RESULT 10  
US-09-964-956-17  
; Sequence 17, Application US/09964956  
; Patent No. 6875570  
; GENERAL INFORMATION:  
; APPLICANT: Gerlach, Valerie L  
; APPLICANT: MacDougall, John R  
; APPLICANT: Smithson, Glenda

;; APPLICANT: Millet, Isabelle  
;; APPLICANT: Stone, David  
;; APPLICANT: Gunther, Erik  
;; APPLICANT: Ellerman, Karen  
;; APPLICANT: Grosse, William M  
;; APPLICANT: Alsobrook II, John P  
;; APPLICANT: Lepley, Denise M  
;; APPLICANT: Bugess, Catherine E  
;; APPLICANT: Padigaru, Muralidhara  
;; APPLICANT: Kekuda, Ramesh  
;; APPLICANT: Spytek, Kimberly A  
;; APPLICANT: Leach, Martin D  
;; APPLICANT: Shinkets, Richard A  
;; TITLE OF INVENTION: No. 6875570el Proteins and Nucleic Acids Encoding Same  
;; FILE REFERENCE: 21402-124  
;; CURRENT APPLICATION NUMBER: US/09/964,956  
;; CURRENT FILING DATE: 2001-09-26  
;; PRIOR APPLICATION NUMBER: 60/235,631  
;; PRIOR FILING DATE: 2000-09-27  
;; PRIOR APPLICATION NUMBER: 60/235,633  
;; PRIOR FILING DATE: 2000-09-27  
;; PRIOR APPLICATION NUMBER: 60/235,808  
;; PRIOR FILING DATE: 2000-09-27  
;; PRIOR APPLICATION NUMBER: 60/236,064  
;; PRIOR FILING DATE: 2000-09-27  
;; PRIOR APPLICATION NUMBER: 60/236,065  
;; PRIOR FILING DATE: 2000-09-27  
;; PRIOR APPLICATION NUMBER: 60/236,066  
;; PRIOR FILING DATE: 2000-09-27  
;; PRIOR APPLICATION NUMBER: 60/236,135  
;; PRIOR FILING DATE: 2000-09-28  
;; PRIOR APPLICATION NUMBER: 60/237,434  
;; PRIOR FILING DATE: 2000-10-03  
;; PRIOR APPLICATION NUMBER: 60/238,321  
;; PRIOR FILING DATE: 2000-10-05  
;; PRIOR APPLICATION NUMBER: 60/238,399  
;; PRIOR FILING DATE: 2000-10-06  
;; PRIOR APPLICATION NUMBER: 60/276,667  
;; PRIOR FILING DATE: 2001-03-16  
;; PRIOR APPLICATION NUMBER: 60/294,823  
;; PRIOR FILING DATE: 2001-05-31  
;; PRIOR APPLICATION NUMBER: 60/304,868  
;; PRIOR FILING DATE: 2001-07-12  
;; NUMBER OF SEQ ID NOS: 127  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 17  
;; LENGTH: 879  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-09-964-956-17

Query Match 7.8%; Score 145.5; DB 2; Length 879;  
Best Local Similarity 23.5%; Pred. No. 1.5e-06;  
Matches 68; Conservative 49; Mismatches 127; Indels 45; Gaps 10;  
  
QY 8 GCRGLSKYRILCDKAEANGVLETATAGVVTSAFMTLTPILVCKVQDSNR----- 61  
DB 561 GCYN-LPEDYIRWED---AWAIGPVTIACLG-----FMCTCIVITVFIKHNTPLVKAS 610  
  
QY 62 -RKMLPTQFLFLGLVIGLTFAPFIIGLDGSTGPTFRFLFGILFSCIFSCLLAHAVSLT 120  
DB 611 GRELC---YILLFGVSLSYCMTEFFFAKPSVICALRLRLGLGTSFAICYSALLTKTNCA 667  
  
QY 121 KLVGRGK-----PLSLVILGLAVGFSLVQDVIAIEYIVLTMTNRTNVVFSE--- 167  
DB 668 RIFDGVNGAQRPKFISPSQVFI---CLGLILVQIVMVSVMLILETPGTRRYTLPEKRE 724  
  
QY 168 ---LSAPRNEDFVLLTY--VLFMLALTFLMSSFTFCGSGFTGKRGHGAHIYLTMLLSIA 222  
DB 725 TVILKCNVKDSSMLISLTYDVVLVILCTVYAFKTRKCPENFNEAKFIGFTMYTTCI--- 780





QY 62 -RKMLPTQFLGLGVLGIFGLTFAFIIGLDGSGTGRFFLFGILFSCFCLLAHAVSLT 120  
Db 611 GRELC---YILFGVGLSYCMWTFFFIAKPSVICALRRGLGSSFAICYSALLTKTNCIA 667  
QY 121 KLVGRK-----PLSLVLGLAVGFSLVQDVIAIEYIVLTWNRNVMVFSE--- 167  
Db 668 RIFDGVKNGAQRKPFISPSQVFI---CLGLILVQIVMVSVMLILEAPGRRYTLAEKRE 724  
QY 168 ---LSAPRRNEDEVLLLTYY--VLFLMALTFMLSSFTFCGSGFTGKRGHAIYILTMLLSIA 222  
Db 725 TVILKCNVDSSMLISLTVDVILCTVYAFKRCPCPNFNKAFIGFTMYTTCI---- 780  
QY 223 IWAMITLLMLPDRRWDITLSSALAANGWVFLLAYVSPFEMLLTKQ 271  
Db 781 IWLAFPIFYVTSSDYRVQTTMCISVLSLGSFVVLGCLFAPKVHILFQ 829

## RESULT 13

US-08-486-270-6  
; Sequence 6, Application US/08486270  
; Patent No. 5807689  
; GENERAL INFORMATION:  
; APPLICANT: Daggett, Lorrie  
; APPLICANT: Ellis, Steven B.  
; APPLICANT: Liaw, Chen  
; APPLICANT: Pontsler, Aaron  
; APPLICANT: Johnson, Edwin C.  
; APPLICANT: Hees, Stephen D.  
; TITLE OF INVENTION: HUMAN METABOTROPIC GLUTAMATE RECEPTORS,  
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SAME AND USES THEREOF  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark  
; STREET: 444 South Flower Street, Suite 2000  
; CITY: Los Angeles  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 90071  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/486,270  
; FILING DATE: 02-JUN-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/072,574  
; FILING DATE: 04-JUN-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Reiter, Stephen E.  
; REGISTRATION NUMBER: 31,192  
; REFERENCE/DOCKET NUMBER: FP41 9772  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619-546-4737  
; TELEFAX: 619-546-9392  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 879 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-486-270-6

Query Match 7.5%; Score 139.5; DB 1; Length 879;  
Best Local Similarity 22.8%; Pred. No. 6.6e-06;  
Matches 66; Conservative 52; Mismatches 126; Indels 45; Gaps 10;  
QY 8 GCRNGLKSKYRLCDKAEAWGIVLETATAGVVTSVAFMLTLPILVCKVQDSNR----- 61  
Db 561 GCYD-LPEDIYIRWED---AWAIGPVTIACLG-----FMCTQMVVTVFIKHNNTPLVKAS 610

QY 62 -RKMLPTQFLGLGVLGIFGLTFAFIIGLDGSGTGRFFLFGILFSCFCLLAHAVSLT 120  
Db 611 GRELC---YILFGVGLSYCMWTFFFIAKPSVICALRRGLGSSFAICYSALLTKTNCIA 667  
QY 121 KLVGRK-----PLSLVLGLAVGFSLVQDVIAIEYIVLTWNRNVMVFSE--- 167  
Db 668 RIFDGVKNGAQRKPFISPSQVFI---CLGLILVQIVMVSVMLILEAPGRRYTLAEKRE 724  
QY 168 ---LSAPRRNEDEVLLLTYY--VLFLMALTFMLSSFTFCGSGFTGKRGHAIYILTMLLSIA 222  
Db 725 TVILKCNVDSSMLISLTVDVILCTVYAFKRCPCPNFNKAFIGFTMYTTCI---- 780  
QY 223 IWAMITLLMLPDRRWDITLSSALAANGWVFLLAYVSPFEMLLTKQ 271  
Db 781 IWLAFPIFYVTSSDYRVQTTMCISVLSLGSFVVLGCLFAPKVHILFQ 829

## RESULT 14

US-08-367-264-6  
; Sequence 6, Application US/08367264  
; Patent No. 6001581  
; GENERAL INFORMATION:  
; APPLICANT: Daggett, Lorrie  
; APPLICANT: Ellis, Steven B.  
; APPLICANT: Liaw, Chen  
; APPLICANT: Pontsler, Aaron  
; APPLICANT: Johnson, Edwin C.  
; APPLICANT: Hees, Stephen D.  
; TITLE OF INVENTION: HUMAN METABOTROPIC GLUTAMATE RECEPTORS,  
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SAME AND USES THEREOF  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark  
; STREET: 444 South Flower Street, Suite 2000  
; CITY: Los Angeles  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 90071  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/367,264  
; FILING DATE: 02-JUN-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/072,574  
; FILING DATE: 04-JUN-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Reiter, Stephen E.  
; REGISTRATION NUMBER: 31,192  
; REFERENCE/DOCKET NUMBER: FP41 9772  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619-546-4737  
; TELEFAX: 619-546-9392  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 879 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-367-264-6

Query Match 7.5%; Score 139.5; DB 2; Length 879;  
Best Local Similarity 22.8%; Pred. No. 6.6e-06;  
Matches 66; Conservative 52; Mismatches 126; Indels 45; Gaps 10;  
QY 8 GCRNGLKSKYRLCDKAEAWGIVLETATAGVVTSVAFMLTLPILVCKVQDSNR----- 61  
Db 561 GCYD-LPEDIYIRWED---AWAIGPVTIACLG-----FMCTQMVVTVFIKHNNTPLVKAS 610



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2006, 15:02:46 ; Search time 11516 Seconds  
(without alignments)  
11925.832 Million cell updates/sec

Title: US-10-600-816-2  
Perfect score: 2456  
Sequence: 1 ataacagcatgaagtgcggt.....ggcagcaaaaaaaaaa 2456

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 48236798 seqs, 27959665780 residues

Total number of hits satisfying chosen parameters: 96473596

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: gb\_est1:\*  
2: gb\_est3:\*  
3: gb\_est4:\*  
4: gb\_est5:\*  
5: gb\_est6:\*  
6: gb\_hic:\*  
7: gb\_est2:\*  
8: gb\_est7:\*  
9: gb\_est8:\*  
10: gb\_est9:\*  
11: gb\_gse1:\*  
12: gb\_gse2:\*  
13: gb\_gse3:\*  
14: gb\_gse4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1830	74.5	1830	6	CR597125	CR597125 full-length
2	1037.6	42.2	1740	6	CR620367	CR620367 full-length
3	965.8	39.3	1022	1	AL556103	AL556103 AL556103
4	935	38.1	1004	1	AL556722	AL556722 AL556722
5	929.4	37.8	1066	1	AL544370	AL544370 AL544370
6	916.8	37.3	1106	1	AL578517	AL578517 AL578517
7	867.4	35.3	902	3	BU552514	BU552514 AGENCOURT
8	864	35.2	1040	3	BM921746	BM921746 AGENCOURT
9	855.2	34.8	983	3	BQ558916	BQ558916 AGENCOURT
10	851.6	34.7	988	3	BM921595	BM921595 AGENCOURT
11	851.2	34.7	862	1	AL571607	AL571607 AL571607
12	843.8	34.4	899	3	BQ936508	BQ936508 AGENCOURT
13	834.8	34.0	897	2	BI820695	BI820695 603034482
14	829.8	33.8	916	5	CF780868	CF780868 AGENCOURT
15	820	33.4	839	1	AL546138	AL546138 AL546138
16	819.8	33.3	918	4	BI372160	BI372160 BX372160
17	817.8	33.3	866	2	BI765773	BI765773 603046537
18	815.4	33.2	1023	3	BQ067433	BQ067433 AGENCOURT
19	808.2	32.9	885	3	BQ430279	BQ430279 AGENCOURT

20	803.4	32.7	899	3	BQ958341	BQ958341 AGENCOURT
21	796	32.4	796	2	BG30281	BG30281 602764675
22	789.8	32.2	812	2	BI824631	BI824631 603033572
23	786	32.0	945	2	BG752462	BG752462 602730954
24	781.2	31.8	865	4	CA454746	CA454746 AGENCOURT
25	779.2	31.7	940	3	BQ685729	BQ685729 AGENCOURT
26	777.4	31.7	933	2	BG823401	BG823401 602726823
27	774.8	31.5	804	2	BM046252	BM046252 603265039
28	771.8	31.4	920	3	BQ920279	BQ920279 AGENCOURT
29	770.4	31.4	895	3	BUI50237	BUI50237 AGENCOURT
30	769.8	31.3	1091	3	BQ922898	BQ922898 AGENCOURT
31	766	31.2	1137	2	BM545793	BM545793 AGENCOURT
32	758.4	30.9	773	1	AL709960	AL709960 DKFZp686D
33	755.8	30.8	962	4	BX370558	BX370558 BX370558
34	752.2	30.6	758	2	BM049765	BM049765 603624288
35	749.4	30.5	795	2	BM049939	BM049939 603624430
36	743.2	30.3	908	2	BG824529	BG824529 602728414
37	723.8	29.5	923	3	BU543952	BU543952 AGENCOURT
38	723.4	29.4	737	8	CN348542	CN348542 170006003
39	722	29.4	880	3	BQ944973	BQ944973 AGENCOURT
40	710.2	28.9	810	2	BI767478	BI767478 603057344
41	708	28.8	781	9	DA647581	DA647581 DA647581
42	707.2	28.8	717	9	DA647448	DA647448 DA647448
43	697	28.4	709	3	BM784295	BM784295 K-EST0062
44	696.8	28.4	788	2	BG118532	BG118532 602348467
45	683.6	27.8	702	2	BG251131	BG251131 602364932

## ALIGNMENTS

RESULT 1	CR597125	1830 bp	mRNA	linear	HTC 21-JUL-2004
CR597125	full-length cDNA clone CSODK011YF17 of HeLa cells	Cot 25-normalized			
LOCUS	of Homo sapiens (human)				
DEFINITION	CR597125				
ACCESSION	CR597125.1	GI:50477932			
VERSION	HTC; CNSLT_cDNA.				
KEYWORDS	Homo sapiens (human)				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 1830)				
AUTHORS	Li, W.B., Gruber, C., Jessee, J. and Polayes, D.				
TITLE	Full-length cDNA libraries and normalization				
JOURNAL	Unpublished				
REMARK	Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Faraday Avenue				
REFERENCE	2 (bases 1 to 1830)				
AUTHORS	Genoscope.				
TITLE	Direct Submission				
JOURNAL	Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE [E-mail : seqref@genoscope.cns.fr]				
COMMENT	- Web : www.genoscope.cns.fr 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.				
FEATURES	Location/Qualifiers				
source	1..1830				
	/organism="Homo sapiens"				
	/mol_type="mRNA"				
	/db_xref="taxon:9606"				
	/clone="CSODK011YF17"				
	/tissue_type="HeLa cells Cot 25-normalized"				
	/plasmid="pCMVSPORT_6"				
ORIGIN					
Query Match	74.5%	Score	1830;	DB 6;	Length 1830;



http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
 Paraday Avenue  
 2 (bases 1 to 1740)  
 Genoscope.  
 Direct Submission  
 Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :  
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
 - Web : www.genoscope.cns.fr)  
 COMMENT  
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
 end enriched, double-strand cDNA was digested with Not I and cloned  
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
 was normalized. Library was constructed by Life Technologies, a  
 division of Invitrogen.

FEATURES  
 Location/Qualifiers  
 1..1740  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0D1018YP20"  
 /tissue type="Placenta Cot 25-normalized"  
 /plasmid="pCMVSPORT\_6"

ORIGIN  
 Query Match 42.2%; Score 1037.6; DB 6; Length 1740;  
 Best Local Similarity 99.1%; Pred. No. 1.5e-209;  
 Matches 1043; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 141 GCGGGATAGCTGTCCAAAGGTCTCCCCAGCACTGAGAGCTCGCTGTGCCCTCTTGC 200  
 DB 1 GCGGGATAGCTGTCCAAAGGTCTCCCCAGCACTGAGAGCTCGCTGTGCCCTCTTGC 60  
 QY 201 GCGCGGAGACGACCAAGATTACGCGCAACGCTTGGCACTAGGCTCCAGAAAGGCTA 260  
 DB 61 GCGCGGAGACGACCAAGATTACGCGCAACGCTTGGCACTAGGCTCCAGAAAGGCTA 120  
 QY 261 CAACAGTCCCTGATGTTGCCGCAATGGCTTGAATCCAAAGTACTACAGACTTTGTGATA 320  
 DB 121 CAACAGTCCCTGATGTTGCCGCAATGGCTTGAATCCAAAGTACTACAGACTTTGTGATA 180  
 QY 321 AGCTGAAGCTTGGGCGATCGCTAGAAACGGTGGCCACAGCGGGGTGTGACCTCGG 380  
 DB 181 AGCTGAAGCTTGGGCGATCGCTAGAAACGGTGGCCACAGCGGGGTGTGACCTCGG 240  
 QY 381 TGGCTTCACTGCTCACTCTCCGATCTCGTCTGCAAGTGCAGGACTCCACAGCGCAA 440  
 DB 241 TGGCTTCACTGCTCACTCTCCGATCTCGTCTGCAAGTGCAGGACTCCACAGCGCAA 300  
 QY 441 AAATGCTGCTACTCAGTTTCTCTCTCTCTGGGTGTGTGGGCATCTTTGGCTCACCT 500  
 DB 301 AAATGCTGCTACTCAGTTTCTCTCTCTCTGGGTGTGTGGGCATCTTTGGCTCACCT 360  
 QY 501 TGGCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 560  
 DB 361 TGGCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420  
 QY 561 TGGCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 620  
 DB 421 TGGCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480  
 QY 621 TGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 680  
 DB 481 TGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 540  
 QY 681 TAGTCAGAGATGTTATCGCTATTGAATATATATATATATATATATATATATATATATAT 740  
 DB 541 TAGTCAGAGATGTTATCGCTATTGAATATATATATATATATATATATATATATATATAT 600  
 QY 741 ATGCTCTTTCAGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTC 800  
 DB 601 ATGCTCTTTCAGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTC 660  
 QY 801 AGTCTCTCTTCTTGTAGTGGCGCTGACCTCTCATGTCTCTCTCTCTCTCTCTCTCTCTCT 860

DB 661 ACGTCTCTTCTTGTATGGGCTGACCTTCTCATGTCTCTCTCTCTCTCTCTCTCTCTCTCT 720  
 QY 861 TCACGGCTGGAAGAGACATGGGGCCACATCTTACCTCAGATGCTCTCTCTCTCTCTCTCT 920  
 DB 721 TCACGGCTGGAAGAGACATGGGGCCACATCTTACCTCAGATGCTCTCTCTCTCTCTCT 780  
 QY 921 TCTGGGTGGCTTGGATCACCTGCTCATGCTTCTCTGACTTTGACCGCAGGTGGATGACA 980  
 DB 781 TCTGGGTGGCTTGGATCACCTGCTCATGCTTCTCTGACTTTGACCGCAGGTGGATGACA 840  
 QY 981 CCATCTCAGCTCCGCTTGGCTGCAATGGCTGGGTGCTCTGTGGCTTATGTAGTC 1040  
 DB 841 CCATCTCAGCTCCGCTTGGCTGCAATGGCTGGGTGCTCTGTGGCTTATGTAGTC 900  
 QY 1041 CCGAGTTTGGCTGCTCACAAGACGAAACCCATGATTCCTGTGGAGTGCTT 1100  
 DB 901 CCGAGTTTGGCTGCTCACAAGACGAAACCCATGATTCCTGTGGAGTGCTT 960  
 QY 1101 TCTGTAACCTCAACTCGTGAAGAGAGCTATGGTGTGGAGAACAGAGCTTCTCTCAAG 1160  
 DB 961 TCTGTAACCTCAACTCGTGAAGAGAGCTATGGTGTGGAGAACAGAGCTTCTCTCAAG 1020  
 QY 1161 AGGAATCACTCAAGGTTTGAAGACAGGG 1192  
 DB 1021 AGGAATCACTCAAGGTTACAGATGAGCTGG 1052

RESULT 3  
 AL556103 1022 bp mRNA linear EST 30-MAR-2004  
 LOCUS AL556103 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens  
 DEFINITION cDNA clone CS0DK011YF17 5-PRIME, mRNA sequence.  
 ACCESSION AL556103  
 VERSION AL556103.3 GI:45860822  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 1022)  
 AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished (2001)  
 COMMENT On Feb 15, 2001 this sequence version replaced gi:31277907.  
 Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 2 rue Gaston Cremieux, Cp 5706 - 91057 EVRY cedex - FRANCE  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
 end enriched, double-strand cDNA was digested with Not I and cloned  
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
 was normalized. Library was constructed by Life Technologies, a  
 division of Invitrogen. This sequence belongs to sequence cluster  
 6601.f  
 For more information about this cluster, see  
 http://www.genoscope.cns.fr/cdna?s=CS0DK011CC09QP14c=6601.f.  
 Location/Qualifiers  
 1..1022  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0DK011YF17"  
 /cell\_type="HELA CELLS COT 25-NORMALIZED"  
 /cell\_line="HELA"  
 /clone lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-oligo(dT)  
 primer. Five prime end enriched, double-strand cDNA was  
 digested with Not I and cloned into the Not I and EcoR V  
 sites of the pCMVSPORT 6 vector. Library was normalized."

FEATURES  
 source  
 1..1022

ORIGIN  
 Query Match 39.3%; Score 965.8; DB 1; Length 1022;  
 |||||

[illegible]

```
Db 361 ATCGGACTGGAGCGGAGCAGAGGCCCCACAGCGCTCTCTCTCTTTGGGATCCTCTTTTC 420
Qy 572 ATCTGCTTCTCTCCCTGCTGGCTCATGCTGTGAGTCTGACCAAGCTGTCGGGGGAGG 631
Db 421 ATCTGCTTCTCTCCCTGCTGGCTCATGCTGTGAGTCTGACCAAGCTGTCGGGGGAGG 480
Qy 632 AAGCCCTTCTCCCTGTTGGTGATTTCTGGGTCTGGCGTGGGCTTACAGCTAGTCCAGGAT 691
Db 481 AAGCCCTTCTCCCTGTTGGTGATTTCTGGGTCTGGCGTGGGCTTACAGCTAGTCCAGGAT 540
Qy 692 GTTATCGCTATTGAATATATGTCCTGACCATGAATAGGACCAACGTCATGTCCTTTTCT 751
Db 541 GTTATCGCTATTGAATATATGTCCTGACCATGAATAGGACCAACGTCATGTCCTTTTCT 600
Qy 752 GAGCTTTCGGCTCTCTGTCGCAATGAAGACTTTGCTCTGCTGCTACCTACGTCCTCTTC 811
Db 601 GAGCTTTCGGCTCTCTGTCGCAATGAAGACTTTGCTCTGCTGCTACCTACGTCCTCTTC 660
Qy 812 TTGATGGCGCTGACCTTCTCTATGTCCTCTTCCCTTCTGCTGCTTCCAGGGCTGG 871
Db 661 TTGATGGCGCTGACCTTCTCTATGTCCTCTTCCCTTCTGCTGCTTCCAGGGCTGG 720
Qy 872 AAGAGACATGGGGCCACATCTACCTCAGCATGCTCTCTCCATTCGCCATCTGGGTGGCC 931
Db 721 AAGAGACATGGGGCCACATCTACCTCAGCATGCTCTCTCCATTCGCCATCTGGGTGGCC 780
Qy 932 TGATCACCCTGCTCATGCTTCTGACTTTGACCGGAGGTGGATGACACCATCTCAGC 991
Db 781 TGATCACCCTGCTCATGCTTCTGACTTTGACCGGAGGTGGATGACACCATCTCAGC 840
Qy 992 TCCGCTTGGCTGCCAATGGCTGGGTTCCTGCTTGGCTTATGTTAGTCCGAGTTTGG 1051
Db 841 TCCGCTTGGCTGCCAATGGCTGGGTTCCTGCTTGGCTTATGTTAGTCCGAGTTTGG 900
Qy 1052 CTGCTCACAAGCAACCAACCCATGATTAATCTGTTGAGGATGCTTTCTGTAACCT 1111
Db 901 CTGCTCACAAGCAACCAACCCATGATTAATCTGTTGAGGATGCTTTCTGTAACCTCA 959
Qy 1112 CAATCTGCTGAAGAGACTATGGT 1136
Db 960 MTGCTGAAGAAGASTATGTGGAG 984

RESULT 5
LOCUS AL544370
DEFINITION AL544370 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
clone CS0D1018YP20 5-PRIME, mRNA sequence.
ACCESSION AL544370
VERSION AL544370.3 GI:45744875
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 1066)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
On Feb 15, 2001 this sequence version replaced gi:31266214.
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
6601.f
For more information about this cluster, see
```

```
http://www.genoscope.cns.fr/cdna?s=CS0D1018DH10QP1&c=6601.f.
Location/Qualifiers
1. 1066
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1018YP20"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 37.8%; Score 929.4; DB 1; Length 1066;
Best Local Similarity 97.7%; Pred. No. 1.4e-186;
Matches 958; Conservative 5; Mismatches 16; Indels 2; Gaps 2;

Qy 166 CCAGCAGCTGAGGAGCTCGCCTGCTGCCCTTTGCGCGGGGAGCAGCAGCAAGTTCAC 225
Db 22 CCGGGATTGAGGAGCTCGCCTGCTGCCCTTTGCGCGGGGAGCAGCAGCAAGTTCAC 81
Qy 226 GGCACACGCTTGGCAGCTAGGCTCCAGAAATGGTACAAAGTCCCTGATGTTGCCGCAA 285
Db 82 GGCACACGCTTGGCAGCTAGGCTCCAGAAATGGTACAAAGTCCCTGATGTTGCCGCAA 141
Qy 286 TGGCTGAAATCAAGTACTACAGACTTTGTGATAAGGCTGAAGCTTGGGGCATCGTCT 345
Db 142 TGGCTGAAATCAAGTACTACAGACTTTGTGATAAGGCTGAAGCTTGGGGCATCGTCT 201
Qy 346 AGAAGCGGTGGCCACAGCGGGTGTGACCTCGGTGGGCTTCATGCTCACTCTCCGAT 405
Db 202 AGAAGCGGTGGCCACAGCTGGGGTGTGACCTCGGTGGGCTTCATGCTCACTCTCCGAT 261
Qy 406 CTTCTGCTCAAGGTGTCAGGACTCCAAAGCGGAAAAATGCTGCTACTCAGTTCCTTT 465
Db 262 CTTCTGCTCAAGGTGTCAGGACTCCAAAGCGGAAAAATGCTGCTACTCAGTTCCTTT 321
Qy 466 CTTCTGCGGTGTTGGGGATCTTTGGGCTCACTTTCGCTTCATCATCGGATGAGCGG 525
Db 322 CTTCTGCGGTGTTGGGGATCTTTGGGCTCACTTTCGCTTCATCATCGGATGAGCGG 381
Qy 526 GAGCAGAGGCGCCACAGCGCTTCTCTCTTTGGGATCTCTTTTCCATCTGCTTCCTG 585
Db 382 GAGCAGAGGCGCCACAGCGCTTCTCTCTTTGGGATCTCTTTTCCATCTGCTTCCTG 441
Qy 586 CTTCTGCGGTGTTGGGGATCTTTGGGCTCACTTTCGCTTCATCATCGGATGAGCGG 645
Db 442 CTTCTGCGGTGTTGGGGATCTTTGGGCTCACTTTCGCTTCATCATCGGATGAGCGG 501
Qy 646 GTTGTGATTTGGGTCTGGCGGTTCAGCTTTCAGCTTTCAGGATGTTATCGCTATTGA 705
Db 502 GTTGTGATTTGGGTCTGGCGGTTCAGCTTTCAGCTTTCAGGATGTTATCGCTATTGA 561
Qy 706 ATATATGTCCTGACCAATAGGACCAACGTCATGCTTTCTGAGCTTTCGCTCC 765
Db 562 ATATATGTCCTGACCAATAGGACCAACGTCATGCTTTCTGAGCTTTCGCTCC 621
Qy 766 TCGTCCGAATGAAGACTTTGCTCTCTGCTCACTACGTCCTCTTCTGATGGCGTGAC 825
Db 622 TCGTCCGAATGAAGACTTTGCTCTCTGCTCACTACGTCCTCTTCTGATGGCGTGAC 681
Qy 826 CTTCTCATGTCCTCTCTTCACTTCTGCTGTTTCCCTTACGGGCTGGAAGACATGGGC 885
Db 682 CTTCTCATGTCCTCTCTTCACTTCTGCTGTTTCCCTTACGGGCTGGAAGACATGGGC 741
Qy 886 CCACATCTACCTCAGATGCTCTCTCCATTCGCAATCGGCTGGCTGATCACCCTGCT 945
Db 742 CCACATCTACCTCAGATGCTCTCTCCATTCGCAATCGGCTGGCTGATCACCCTGCT 801
Qy 946 CATGCTTCTGACTTTGACCGCAGGTGGGATGACACCATCTCAGCTCCGCTTGGCTGC 1005
```

Db	802	CATGCTCTCTGACTTTTGACCGCAGTGGGATGACACCATCTCTCAGCTCGCCTTGGCTGC	861
Qy	1006	CAATGGCTGGGTGTTCCCTGTTGGCTATGTTAGTCCCGAGTTTGGCTGCTCACAAAGCA	1065
Db	862	CAATGGCTGGGTGTTCCCTGTTGGCTATGTTAGTCCCGAGTTTGGCTGCTCACAAAGCA	921
Qy	1066	ACGAAACCCATGGATATATCTGTTGGATGATCTTCTGTTAAACCTCAACCTCGTGAAGAA	1125
Db	922	ACGAAA-CCCATKGAATATCTGTTGGATGATCTTCTGTTAA-CTCAACTCGTGAAGAG	979
Qy	1126	GAGCTATGTTGGGAGACAG	1146
Db	980	RAGTATTGTTKGAGAAAGAG	1000
RESULT 6			
AL578517/c			
LOCUS	AL578517	1106 bp	mRNA linear EST 07-APR-2004
DEFINITION	AL578517 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens		
ACCESSION	CDNA clone CS0DK011YF17 3-PRIME, mRNA sequence.		
VERSION	AL578517		
KEYWORDS	EST.		
SOURCE	AL578517.3 GI:46257411		
ORGANISM	Homo sapiens (human)		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
TITLE	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;		
JOURNAL	Hominidae; Homo.		
COMMENT	1 (bases 1 to 1106)		
	Li W.B., Gruber C., Jessee J. and Polayes D.		
	Full-length cDNA libraries and normalization		
	Unpublished (2001)		
	On Feb 16, 2001 this sequence version replaced gi:31316729.		
	Contact: Genoscope		
	Genoscope - Centre National de Sequencage		
	2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE		
	Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr		
	1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime		
	end enriched, double-strand cDNA was digested with Not I and cloned		
	into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library		
	was normalized. Library was constructed by Life Technologies, a		
	division of Invitrogen. This sequence belongs to sequence cluster		
	6601.f		
	For more information about this cluster, see		
	http://www.genoscope.cns.fr/cdna?S=CS0DK011CC09NP1&c=6601.f.		
FEATURES	Location/Qualifiers		
source	1..1106		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="CS0DK011YF17"		
	/cell_type="HELA CELLS COT 25-NORMALIZED"		
	/cell_line="HELA"		
	/clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"		
	/note="1st strand cDNA was primed with a NotI-oligo(dT)		
	primer. Five prime end enriched, double-strand cDNA was		
	digested with Not I and cloned into the Not I and EcoR V		
	sites of the pCMVSPORT 6 vector. Library was normalized."		
ORIGIN			
Query Match	37.3%; Score 916.8; DB 1; Length 1106;		
Best Local Similarity	98.1%; Pred. No. 6.8e-184;		
Matches 935;	Conservative 10; Mismatches 6; Indels 2; Gaps 2;		
Qy	1418	AAACTGTACAGACACTAGCGGAAGTTTGCTCCCTCCAGCCTCAACCAATCTT	1477
Db	952	AAATKTACAGACACTAGCGGAACA-TTTKCTCCCTCCAGCCTCAACCAATCTT	894
Qy	1478	CCATGCTGGGCTGATGTGGGTAGTAGACATCCAGTCTTGTAGAGCGCTGTAGTA-TTT	1536
Db	893	CCATTSTGGGCTGATGTGGGTAGTAGACATCCAGTCTTGTAGAGCGCTGTAGTATTTT	834
Qy	1537	TTTTTTTTTGTCATCTCTTTGGATACTTCTTTAAAGTGGGAGTCTCAGGCAACTCAAG	1596
Db	833	TTTTTTTTTGTCATCTCTTTGGATACTTCTTTAAAGTGGGAGTCTCAGGCAACTCAAG	774
Qy	1597	TTTAGACCTTACTCTTTTGTGTTTGTGTTTGAACAGAGATCTTGTCTGTCAACCCAGCG	1656
Db	773	TTTAGACCTTACTCTTTTGTGTTTGTGTTTGAACAGAGATCTTGTCTGTCAACCCAGCG	714
Qy	1657	TTGAGTGAGTGGTGGATACAGCCAGTGAGCCTCGACCACTGTGTCTCAAGAATC	1716
Db	713	TTGAGTGAGTGGTGGATACAGCCAGTGAGCCTCGACCACTGTGTCTCAAGAATC	654
Qy	1717	CTCCCATCTCCATCTCCCAAGTGTGGATGACAGGCGTGAGCCACAGCTCCAGCCTA	1776
Db	653	CTCCCATCTCCATCTCCCAAGTGTGGATGACAGGCGTGAGCCACAGCTCCAGCCTA	594
Qy	1777	GGCCCTTAATCTTGTGTTTATTTTCCATGGGATAAAGTCTCGTCACTGTGAGTCAACGCT	1836
Db	593	GGCCCTTAATCTTGTGTTTATTTTCCATGGGATAAAGTCTCGTCACTGTGAGTCAACGCT	534
Qy	1837	GGCTCACACAGCTCTAGGGGCTGCTCTCTAACTCACAGTGGGTTTGTGAGGCTCTGT	1896
Db	533	GGCTCACACAGCTCTAGGGGCTGCTCTCTAACTCACAGTGGGTTTGTGAGGCTCTGT	474
Qy	1897	GGCCAGAGCAGCTGCTGATATCTGAGCAAAATAGCAAAAGCCTCTCTCAGCCCACTG	1956
Db	473	GGCCAGAGCAGCTGCTGATATCTGAGCAAAATAGCAAAAGCCTCTCTCAGCCCACTG	414
Qy	1957	CCTGAATCTACACTGGAAGCCAACTTGTGGCACCCCGCTCCCAACCCCTTCTTGCCTG	2016
Db	413	CCTGAATCTACACTGGAAGCCAACTTGTGGCACCCCGCTCCCAACCCCTTCTTGCCTG	354
Qy	2017	GGTAGGAGGCTTAAGATCACCTAAATTTACTCATCTCTTAGTGTGCTCAGATTG	2076
Db	353	GGTAGGAGGCTTAAGATCACCTAAATTTACTCATCTCTTAGTGTGCTCAGATTG	294
Qy	2137	AAACTTGTCTCAATTCGAGATCTAATCTCCCTCCCTAGCTCTGCAGGAATTTTTCAG	2196
Db	233	AAACTTGTCTCAATTCGAGATCTAATCTCCCTCCCTAGCTCTGCAGGAATTTTTCAG	174
Qy	2197	ACCTCAGTACAGCAAGCCGGTGTCTTGTGACAGAAATTTGTAGATCATCTCATT	2256
Db	173	ACCTCAGTACAGCAAGCCGGTGTCTTGTGACAGAAATTTGTAGATCATCTCATT	114
Qy	2257	CAAAATCTCTGGGCTGATACTTCTCTCATCTTGCACCCCAACCTCTGTAATAGATTAC	2316
Db	113	CAAAATCTCTGGGCTGATACTTCTCTCATCTTGCACCCCAACCTCTGTAATAGATTAC	54
Qy	2317	CGCATTTACGGCTGATCTTGTAAAGTGGGATGTTCTCTCTTAATAGGAGGAGTGT	2369
Db	53	CGCATTTACGGCTGATCTTGTAAAGTGGGATGTTCTCTCTTAATAGGAGGAGTGT	1
RESULT 7			
BU552514			
LOCUS	BU552514		
DEFINITION	AGENCOURT_10333192 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6576305		
ACCESSION	5', mRNA sequence.		
VERSION	BU552514		
KEYWORDS	BU552514.1 GI:22902786		
SOURCE	EST.		
ORGANISM	Homo sapiens (human)		
	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;		
	Hominidae; Homo.		
	1 (bases 1 to 902)		
	NIH-MGC http://mgc.nci.nih.gov/.		
	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished (1999)		



## COMMENT

Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs@mail.nih.gov](mailto:cgapbs@mail.nih.gov)  
Tissue Procurement: DCTD/DTP  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNLN at:

Plate: LLCM2774 row: j column: 17  
High quality sequence stop: 716.

## FEATURES

```

1. .902
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6576305"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 40"
/note="Organ: prostate; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH MGC Library."

```

## ORIGIN

Query Match	35.3%	Score 867.4;	DB 3;	Length 902;
Best Local Similarity	98.4%	Pred. No. 2.1e-173;		
Matches 887;	Conservative	0;	Mismatches 11;	Indels 3;
				Gaps 1;

164	Qy	CCCCAGCACTGAGGAGCTCGCTGCTGCCCCTCTTTGGCGGGGAAGCAGCAACCAAGTTC	223
1	Db	CCCCAGCACTGAGGAGCTCGCTGCTGCCCCTCTTTGGCGGGGAAGCAGCAACCAAGTTC	60
224	Qy	ACGGCCAAACGCCCTTGGCACTAGGCTCAGAAATGGCTACAACAGTCCCTGATGGTTGCCGC	283
61	Db	ACGGCCAAACGCCCTTGGCACTAGGCTCAGAAATGGCTACAACAGTCCCTGATGGTTGCCGC	120
284	Qy	AATGGCCTGAAATCCAAAGTACTACAGACTTTGTGATAAGGCTGAAGCTTTGGGGCATCGTC	343
121	Db	AATGGCCTGAAATCCAAAGTACTACAGACTTTGTGATAAGGCTGAAGCTTTGGGGCATCGTC	180
344	Qy	CTAGAAACCGTGGCCACAGCGGGGTTGTGACCTTCGCTGGGCTTTCATGCTCATCTCCCG	403
181	Db	CTAGAAACCGTGGCCACAGCGGGGTTGTGACCTTCGCTGGGCTTTCATGCTCATCTCCCG	240
404	Qy	ATCCTCGTCTGCAAGGTGCAGGACTCAACAGGGCGAAATGCTGCCCTACTCAGTTTCTC	463
241	Db	ATCCTCGTCTGCAAGGTGCAGGACTCAACAGGGCGAAATGCTGCCCTACTCAGTTTCTC	300
464	Qy	TTCTCTCTGGGTGTGTTGGGCATCTTTGGGCTCACCTTTGGGCTTTCATCATCGGACTGGAC	523
301	Db	TTCTCTCTGGGTGTGTTGGGCACTTTGGGCTCACCTTTGGGCTTTCATCATCGGACTGGAC	360
524	Qy	GGGAGCACAGGGGCCACACGGTCTTCCCTTTTGGGATCCCTCTTTTCCATCTGCTTCTCC	583
361	Db	GGGAGCACAGGGGCCACACGGTCTTCCCTCTTTTGGGATCCCTCTTTTCCATCTGCTTCTCC	420
584	Qy	TGCTGCTGCTCATGCTGTCAGTCTCACCAAGCTCGTCCGGGGGAGGAAGCCCTTTTCC	643
421	Db	TGCCCTGCTGGCTCATGCTGTCAAGTCTGACCAAGCTCGTCCGGGGGAGGAAGCCCTTTTCC	480
644	Qy	CTGTTGCTGATTTCTGGGCTCTGGGCTTCAGCCCTAGTCCAGGATGTTATCGCTATT	703
481	Db	CTGTTGCTGATTTCTGGGCTCTGGGCTTCAGCCCTAGTCCAGGATGTTATCGCTATT	540
704	Qy	GAATATATTGCTGACCAATGAAATAGGACCAAGTCAATGCTTTTCTGAGCTTTCCGCT	763
541	Db	GAATATATTGCTGACCAATGAAATAGGACCAAGTCAATGCTTTTCTGAGCTTTCCGCT	600

## ORIGIN

AGENCY	AGENCOURT_10029407 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6481198
DEFINITION	5', mRNA sequence.
ACCESSION	BQ958916
VERSION	BQ958916.1 GI:22374394
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: <a href="mailto:cgapbs-remail.nih.gov">cgapbs-remail.nih.gov</a> Tissue Procurement: DCTD/DTP
FEATURES	cDNA Library Preparation: Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Plate: LLCM2664 row: d column: 15 High quality sequence stop: 615.
Source	1. 983 Location/Qualifiers /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:6481982" /tissue_type="carcinoma, cell line" /lab_host="DH10B (phage-resistant)" /clone_lib="NIH_MGC_40" /note="Organ: prostate; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
ORIGIN	Query Match 34.8%; Score 855.2; DB 3; Length 983; Best Local Similarity 96.4%; Pred. No. 8.3e-171; Matches 906; Conservative 0; Mismatches 30; Indels 4; Gaps 3;
QY	169 AGCACTGAGGAGCTCGCTGCTGCCCTCTTGGCGGGGAAGCAGACCAAGTTTCACGGC 228
DB	1 AGCACTGAGGAGCTCGCTGCTGCCCTCTTGGCGGGGAAGCAGACCAAGTTTCACGGC 60
QY	229 CAACGCCCTTGGCACTAGGGTCCAGATGGGTCAACAAGTCCCTGATGGTTGCCGCAATGG 288
DB	61 CAACGCCCTTGGCACTAGGGTCCAGATGGGTCAACAAGTCCCTGATGGTTGCCGCAATGG 120
QY	289 CCTGAAATCCAAGTACTACAGCTTTGTGATAGAGCTGAAAGCTTTGGGGCATCGTCTTAG 348
DB	121 CCTGAAATCCAAGTACTACAGCTTTGTGATAGAGCTGAAAGCTTTGGGGCATCGTCTTAG 180
QY	349 AACGCTGCCACACGCGGGTTGACCTCGGTGGCTTCATGCTACTCTCCCGATCCT 408
DB	181 AACGCTGCCACACGCGGGTTGACCTCGGTGGCTTCATGCTACTCTCCCGATCCT 240
QY	409 CGTCTGCAAGTGCAGGACTCCAAACAGCGGAAATAATCGCTACTCAGTTTCTCTTCCT 468
DB	241 CGTCTGCAAGTGCAGGACTCCAAACAGCGGAAATAATCGCTACTCAGTTTCTCTTCCT 300
QY	469 CTTGGGTGTGTGGGCATCTTTGGCTCCTCAGCTCGGCTTCATCATCGGACTGACCGGAG 528
DB	301 CTTGGGTGTGTGGGCATCTTTGGCTCCTCAGCTCGGCTTCATCATCGGACTGACCGGAG 360
QY	529 CACAGGGGCCACAGGCTTTCTTCCTCTTTGGGATCCTCTTTTCCATCTGCTTCTCTGCCT 588

|||||  
361 CACAGGGCCACACGCTTCTCTCTTGGATCCTCTTTTCATCTGCTTCTCCTGCT 420  
|||||  
589 GCTGGCTCATGCTGTGAGTGTACCAAGCTGTGCGGGGAGAAAGCCCTTTCCCTGTT 648  
Db GCTGGCTCATGCTGTGAGTGTACCAAGCTGTGCGGGGAGAAAGCCCTTTCCCTGTT 480  
|||||  
649 GGTGATTCCTGGGTCGCGCGTGGCTTACAGCTAGTCCAGGATGTTATCGCTATTGAATA 708  
Db GGTGATTCCTGGGTCGCGCGTGGCTTACAGCTAGTCCAGGATGTTATCGCTATTGAATA 540  
|||||  
709 TATTGCTCTGACCATGAATAGGACCAACGTCATGCTTTTCTGAGCTTTCCGCTCCTCG 768  
Db TATTGCTCTGACCATGAATAGGACCAACGTCATGCTTTTCTGAGCTTTCCGCTCCTCG 600  
|||||  
769 TGGCAATGAAGACTTTGTCTCTGCTGCTACAGTCTCTCTTGTGATGGCGTGACCTT 828  
Db TGGCAATGAAGACTTTGTCTCTGCTGCTACAGTCTCTCTTGTGATGGCGTGACCTT 660  
|||||  
829 CTTATGCTCTCTCTTCACTTCTGCTGCTTCTTCAAGGCTGCTTCAAGGATGCTGCTGCT 888  
Db CTTATGCTCTCTCTTCACTTCTGCTGCTTCTTCAAGGCTGCTTCAAGGATGCTGCTGCT 720  
|||||  
889 CATCTACCTCAGATGCTCTCTTCCATTCGCTGCTGCTGCTGCTGCTGCTGCTCAT 948  
Db CATCTACCTCAGATGCTCTCTTCCATTCGCTGCTGCTGCTGCTGCTGCTGCTCAT 779  
|||||  
949 GCTTCTGATCTTGACCGAGTGGATGACACCATCTCAGCTCCGCTTGGCTGCGCAA 1008  
Db GCTTCTGATCTTGACCGAGTGGATGACACCATCTCAGCTCCGCTTGGCTGCGCAA 839  
|||||  
1009 TGGCTGGGTGCTCTGCTGCTTATGTT--AGTCCCGAGTTTGGCTGCTCAC-AAAGCA 1065  
Db TGGCTGGGTGCTCTGCTGCTTATGTT--AGTCCCGAGTTTGGCTGCTCAC-AAAGCA 899  
|||||  
1066 ACGAAACCCCTGGATTATCTCTGTTGAGGATGCTTCTGT 1105  
Db ACGAAACCCCTGGATTATCTCTGTTGAGGATGCTTCTGT 939  
  
RESULT 10  
BM921595  
LOCUS BM921595 988 bp mRNA linear EST 12-MAR-2002  
DEFINITION AGENCOURT 6707965 NIH\_MGC\_115 Homo sapiens cDNA clone IMAGE:5753101  
5', mRNA sequence.  
ACCESSION BM921595  
VERSION BM921595.1 GI:19371974  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
1 (bases 1 to 988)  
NIH-MGC <http://imgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgabs-r@mail.nih.gov](mailto:cgabs-r@mail.nih.gov)  
Tissue Procurement: Life Technologies, Inc.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM12788 row: f column: 14  
High quality sequence stop: 694.  
Location/Qualifiers  
1. .988  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"

FEATURES  
source

/clone="IMAGE:5753101"  
/lab\_host="DH10B"  
/clone\_lib="NIH\_MGC\_115"  
/note="Organ: pooled brain, lung, testis; Vector:  
pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA  
source anonymous pool of 6 male brains, age range 23-27; 1  
male lung, age 27; and 1 male testis, age 69. Library is  
oligo-dT primed and directionally cloned (EcoRV site is  
destroyed upon cloning). Average insert size 1.8 kb,  
insert size range 1-3 kb. Library is normalized and  
enriched for full-length clones and was constructed by C.  
Gruber (Invitrogen). Research Genetics tracking code  
021. Note: this is a NIH\_MGC Library."  
  
ORIGIN  
  
Query Match 34.7%; Score 851.6; DB 3; Length 988;  
Best Local Similarity 97.5%; Pred. No. 4.8e-170;  
Matches 906; Conservative 0; Mismatches 17; Indels 6; Gaps 4;  
  
QY 161 TCTCCCCCAGCACTGAGGAGCTCGCTGCTGCCCTCTTGGCGCGGGAAGCAGCACAAG 220  
Db 1 TCTCCCCCAGCACTGAGGAGCTCGCTGCTGCCCTCTTGGCGCGGGAAGCAGCACAAG 60  
QY 221 TTTCAGGGCCAAACGCTTGGCACTAGGGTCCAGAAATGGCTACAAACAGTCCCTCATGTGTC 280  
Db 61 TTTCAGGGCCAAACGCTTGGCACTAGGGTCCAGAAATGGCTACAAACAGTCCCTCATGTGTC 120  
QY 281 CGCAATGGCTGAAATCCAAAGTACTACAGACTTTGTGATAAGGCTGAAAGCTGGGGGATC 340  
Db 121 CGCAATGGCTGAAATCCAAAGTACTACAGACTTTGTGATAAGGCTGAAAGCTGGGGGATC 180  
QY 341 GTCTTAGAAGACGGTGGCCACAGCGGGGTGTGACCTCGGTGGCTTCAATGCTCACTCTC 400  
Db 181 GTCTTAGAAGACGGTGGCCACAGCGGGGTGTGACCTCGGTGGCTTCAATGCTCACTCTC 240  
QY 401 CCGATCCTCGTCTGCAAGGTGCAGGACTCCAAACAGCGGAAATAATGCTCACTAGTTT 460  
Db 241 CCGATCCTCGTCTGCAAGGTGCAGGACTCCAAACAGCGGAAATAATGCTCACTAGTTT 300  
QY 461 CTCCTCTCTCGGTGTGTGGGCATCTTTGGCTCACTCGCTTCATCATCGAGCTG 520  
Db 301 CTCCTCTCTCGGTGTGTGGGCATCTTTGGCTCACCTTCGCTTCACTCGAGCTG 360  
QY 521 GACGGAGCACAGGGCCACACAGCTTCTCTCTTTGGGATCTCTTTTCCATCTCCTTC 580  
Db 361 GACGGAGCACAGGGCCACACAGCTTCTCTCTTTGGGATCTCTTTTCCATCTCCTTC 420  
QY 581 TCCTGCTGCTGGCTCATGCTCTGAGTCTGACCAAGCTCGTCCGGGGAGGAAGCCCTTT 640  
Db 421 TCCTGCTGCTGGCTCATGCTCTGAGTCTGACCAAGCTCGTCCGGGGAGGAAGCCCTTT 480  
QY 641 TCCCTGTGGTGAATCTTGGGTCTGGCGTGGGCTTACAGCTAGTCAGGATGTTATCGCT 700  
Db 481 TCCCTGTGGTGAATCTTGGGTCTGGCGTGGGCTTACAGCTAGTCAGGATGTTATCGCT 540  
QY 701 ATTGAATATATTGCTCTGACCATGAATAGGACCAAGCTCAATGTCTTTTTCAGCTTTCC 760  
Db 541 ATTGAATATATTGCTCTGACCATGAATAGGACCAAGCTCAATGTCTTTTTCAGCTTTCC 600  
QY 761 GCTCTCTGCTGCAATGAAGACTTTGCTCTCTGCTCACCTAGTCTCTTCTTGTATGGGG 820  
Db 601 GCTCTCTGCTGCAATGAAGACTTTGCTCTCTGCTCACCTAGTCTCTTCTTGTATGGGG 660  
QY 821 CTGACCTTCTCATGCTCTCTTCACTTCACTTCTGTGTTCTTTCACGGGCTGGGAAGACAT 880  
Db 661 CTGACCTTCTCATGCTCTCTTCACTTCACTTCTGTGTTCTTTCACGGGCTGGGAAGACAT 720  
QY 881 GGGGGCCACATCTACTCAGGATGCTCTCTCTCCATGCTGCTGCTGGTGGCTGGATCACC 940  
Db 721 GGGGGCCACATCTACTCAGGATGCTCTCTCTCCATGCTGCTGCTGGTGGCTGGATCACC 779  
QY 941 CTGCTCATGCTTCTCTGACTTTTACCGCGAGTGGGATGACACCATCTCAGCTCCGCTTG 1000  
Db CTGCTCATGCTTCTCTGACTTTTACCGCGAGTGGGATGACACCATCTCAGCTCCGCTTG

```
Db      780 CTGCTCATGCTTCCTGACNTGACCGCAGGTGGGATGACACCATCCTCAGCTCCGCTTG 839
Qy      1001 GCTGCAATGGCTGGG-TGTTCCTGTTGGCTTATGTTAGTCCCGAG---TTTGTGCTGCT 1056
Db      840 GCTGCAATGGCTGGGTTGCTCCTGGTGGCTTATGTTTANTCCCGAGGTTTGTGCTGCT 899
Qy      1057 CACAAA-GCAAGAACCCCATGGATTAT 1084
Db      900 CACAAGGCAACGAACCCCTGGGATTT 928

RESULT 11
AL571607/c
LOCUS      AL571607 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
DEFINITION clone CS0D1025YF10 3-PRIME, mRNA sequence.
ACCESSION AL571607
VERSION    AL571607.1 GI:12929072
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo
REFERENCE  1 (bases 1 to 862)
AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE     Full-length cDNA libraries and normalization
JOURNAL   Unpublished (2001)
COMMENT   Contact: Genoscope
            Genoscope - Centre National de Sequencage
            2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
            Email: seqrefgenoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen. This sequence belongs to sequence cluster
            6601.f
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0D1025D05NP1&c=6601.f.

FEATURES             source
            Location/Qualifiers
            1..862
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="CS0D1025YF10"
            /tissue_type="PLACENTA COT 25-NORMALIZED"
            /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
            /note="1st strand cDNA was primed with a NotI-oligo(dT)
            primer. Five prime end enriched, double-strand cDNA was
            digested with Not I and cloned into the Not I and EcoR V
            sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
Query Match      34.7%; Score 851.2; DB 1; Length 862;
Beat Local Similarity 98.6%; Pred. No. 5.8e-170;
Matches 850; Conserved 7; Mismatches 5; Indels 0; Gaps 0;

Qy      1532 TATTTTATTTTGTCTCATCTCTTGATACTTCTTTTAAGTGGAGTCTCAGGCAAC 1591
Db      862 TATTTTATTTTGTCTCATCTCTTGATACTTCTTTTAAGTGGAGTCTCAGGCAAC 803
Qy      1592 TCAAGTTTAGACCCCTACTCTTTTGTGTTTTTGAACAGGATCTTGTCTGTCCACC 1651
Db      802 TCAAGTTTAGACCCCTACTCTTTTGTGTTTTTGAACAGGATCTTGTCTGTCCACC 743
Qy      1652 CAGGCTTGAGTGCAGTGGTGCATCACGCCAGTGCAGCTCGACCACCTGTGCTCAAG 1711
Db      742 CAGGCTTGAGTGCAGTGGTGCATCACGCCAGTGCAGCTCGACCACCTGTGCTCAAG 683
Qy      1712 CAATCTCCCATCTCCATCTCCCAAGTGTCTGGGATGACAGGCGTGAGCCAGAGTCCCA 1771
Db      682 CAATCTCCCATCTCCATCTCCCAAGTGTCTGGGATGACAGGCGTGAGCCAGAGTCCCA 623
```

```
Qy      1772 GCCTAGGCGCTTAATCTTTGCTGTTATTTTCCATGAGCTAAAGGTCTGGTCACTCAGCTC 1831
Db      622 GCCTAGGCGCTTAATCTTTGCTGTTATTTTCCATGAGCTAAAGGTCTGGTCACTCAGCTC 563
Qy      1832 AGCTGGCTCACACAGCTCTAGGGGCTGCTCTTAAGTCACTACAGTGGTGGTGGAGG 1891
Db      562 AGCTGGCTCACACAGCTCTAGGGGCTGCTCTTAAGTCACTACAGTGGTGGTGGAGG 503
Qy      1892 TCTGTGGCCGAGAGCAGACCTGTCATATCTGAGCAAAAATAGCAAAAGCCTCTCTCAGCCC 1951
Db      502 TCTGTGGCCGAGAGCAGACCTGTCATATCTGAGCAAAAATAGCAAAAGCCTCTCTCAGCCC 443
Qy      1952 ACTGCGCTGAATCTACACTGGAAGCAACTTGTCTGGACCCCGCTCCCAACCTCTTT 2011
Db      442 ACTGCGCTGAATCTACACTGGAAGCAACTTGTCTGGACCCCGCTCCCAACCTCTTT 383
Qy      2012 GCCTGGGTAGGAGGCTTAAAGATCACCTAAATTTACTCATCTCTAGTCTGCCTCA 2071
Db      382 GCCTGGGTAGGAGGCTTAAAGATCACCTAAATTTACTCATCTCTAGTCTGCCTCA 323
Qy      2072 CATTTGGGCTCAGCAGCTCCCGCAGCAACAATTCACAGGTCAACCCCTCTCTTTGCACTG 2131
Db      322 CATTTGGGCTCAGCAGCTCCCGCAGCAACAATTCACAGGTCAACCCCTCTCTTTGCACTG 263
Qy      2132 TCCCAAACTTGTCTGTTCAATTTCCGAGATCTAATCTCCCGCTACGCTCTGCCAGGAATCT 2191
Db      262 TCCCAAACTTGTCTGTTCAATTTCCGAGATCTAATCTCCCGCTACGCTCTGCCAGGAATCT 203
Qy      2192 TTCAGACCTCACTAGCACAAGCCGGTTGCTCTTGTGAGGAGAAATTTGTAGATCATTTCT 2251
Db      202 TTCAGACCTCACTAGCACAAGCCGGTTGCTCTTGTGAGGAGAAATTTGTAGATCATTTCT 143
Qy      2252 CACTTCAAATTCCTGGGCTGATCTTCTCATCTTGCACCCCAACCTCTCTGTAATAGA 2311
Db      142 CACTTCAAATTCCTGGGCTGATCTTCTCATCTTGCACCCCAACCTCTCTGTAATAGA 83
Qy      2312 TTTACCGGATTTACGGCTGCAATCTGTAAGTGGGATGCTCTCTTAAGGAGGAGTGTTC 2371
Db      82 TTTACCGGATTTACGGCTGCAATCTGTAAGTGGGATGCTCTCTTAAGGAGGAGTGTTC 23
Qy      2372 ATTGTATTAATAGTTATTTCACC 2393
Db      22 ATTGTATTAATAGTTATTTCACC 1

RESULT 12
BQ936508
LOCUS      BQ936508
DEFINITION AGENCOURT_10058421 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6480086
            5', mRNA sequence.
ACCESSION BQ936508
VERSION    BQ936508.1 GI:22351891
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo
            1 (bases 1 to 899)
            NIH-MGC http://mgc.nci.nih.gov/
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTB/Drp
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2659 row: e column: 15
```



Db 121 CGCAATGCCTGAATCAAGTACTACAGACTTTGTGATAGGCTGAAGCTTGGGGCAT 180  
Qy 340 CQTCTAGAAACGGTGGCCACAGCGGGTGTGACCTCGGTGGCTTCATGCTCACTCT 399  
Db 181 CQTCTAGAAACGGTGGCCACAGCGGGTGTGACCTCGGTGGCTTCATGCTCACTCT 240  
Qy 400 CCCGATCCTCTGCTCAAGGTCAGGACTCCAAAGGCGGAAAAATGCTGCTACTCAGTT 459  
Db 241 CCCGATCCTCTGCTCAAGGTCAGGACTCCAAAGGCGGAAAAATGCTGCTACTCAGTT 300  
Qy 460 TCTCTTCTCTCGGTGTGGGATCTTTGGGCTCAGCTTCGGCTTCATCATCGGACT 519  
Db 301 TCTCTTCTCTCGGTGTGGGATCTTTGGGCTCAGCTTCGGCTTCATCATCGGACT 360  
Qy 520 GGACGGGAGCAGAGGCGCCACAGCTTCTCTCTTTGGGATCCCTTTTCCATCTGCTT 579  
Db 361 GGACGGGAGCAGAGGCGCCACAGCTTCTCTCTTTGGGATCCCTTTTCCATCTGCTT 420  
Qy 580 CTCCTGCTCTGCTCATGCTGTGAGTCTGACCAAGCTCGTCCGGGGAGGAAGCCCT 639  
Db 421 CTCCTGCTCTGCTCATGCTGTGAGTCTGACCAAGCTCGTCCGGGGAGGAAGCCCT 480  
Qy 640 TTCCTGTGTGATCTGGGCTGCGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCT 699  
Db 481 TTCCTGTGTGATCTGGGCTGCGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCT 540  
Qy 700 TATTGAATATATGCTTGCATGACCATGAATAGGACCAAGCTGCTTCTTCTGAGCTTTC 759  
Db 541 TATTGAATATATGCTTGCATGACCATGAATAGGACCAAGCTGCTTCTTCTGAGCTTTC 600  
Qy 760 CGCTCCTCGTCAATGAAGACTTTGCTCTCTGCTCAGCTTCTGCTTCTTCTGATGCG 819  
Db 601 CGCTCCTCGTCAATGAAGACTTTGCTCTCTGCTCAGCTTCTGCTTCTTCTGATGCG 660  
Qy 820 GCTGACCTTCTCATGCTCTCTTCCCTTCCCTTCTGCTTCTGCTTCTGCTTCTGCTTCT 879  
Db 661 GCTGACCTTCTCATGCTCTCTTCCCTTCCCTTCTGCTTCTGCTTCTGCTTCTGCTTCT 720  
Qy 880 TGGGGCCACATCTACCTCAGCATGCTCTCTTCCCTTCCCTTCTGCTTCTGCTTCTGCTTCT 939  
Db 721 TGGGGCCACATCTACCTCAGCATGCTCTCTTCCCTTCCCTTCTGCTTCTGCTTCTGCTTCT 780  
Qy 940 CTTGC-TCATGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCT 998  
Db 781 CTTGCCTCATGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCT 840  
Qy 999 TGGCTGCCAATCGCTGGGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCT 1048  
Db 841 TGGCTGCC-ATGGCTGGGCTG-TCCTGTTGGTATGTTAGTCCCGAGTTT 888

RESULT 14  
CF780868  
LOCUS  
DEFINITION  
AGENCOURT 15939245 NIH\_MGC.219 Homo sapiens cdna clone  
IMAGE:30523569 5', mRNA sequence.  
ACCESSION  
CF780868  
VERSION  
CF780868.1 GI:37740645  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
1 (bases 1 to 916)  
REFERENCE  
NIH-MGC <http://mgi.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
CONTACT: Daniela S. Gerhard, Ph.D.  
Office of Cancer Genomics  
National Cancer Institute / NIH  
Bldg. 31 Rm10A07 Bethesda, MD 20892

Email: c9apbs-r@mail.nih.gov  
Tissue Procurement: James Martin, University of Iowa  
cdna Library Preparation: M. Bento Soares, University of Iowa  
cdna Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: NDAM601 row: 1 column: 10  
High quality sequence stop: 651.  
FEATURES  
Location/Qualifiers  
source  
1..916  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:30523569"  
/issue\_type="Pooled Chondrosarcoma Tumor cells"  
/lab\_host="DH10B (T1 phage-resistant)"  
/clone\_lib="NIH\_MGC\_219"  
/note="Vector: pYX-Asc; Site 1: EcoRI; Site 2: NotI;  
Library is oligo-dr primed and directionally cloned.  
Denatured RNA was size fractionated on a 1% agarose gel.  
First strand cdna synthesis was primed with oligo-dr  
primer containing a Not I site. Double strand cdna was  
size selected according to rna size fraction, ligated with  
EcoR I adaptor, digested with Not I and then cloned  
directionally into pYX-Asc vector. Average insert size  
0.5-1kb. Adaptors 5'(AATTCGGCAGGAG)3' and 5'(d  
(CTCTGCTCCG)3'. 3' Linker sequence - GGGCGCTGAGAGCC T18.  
Sequencing primers 3'end: T3 promoter primer 5'd  
(AATACCTCCTAAAGGA)3'. 5' End: T7 promoter primer 5'd  
(TAATACGACTCATAGG)3'. Library was constructed in the  
laboratory of M. Bento Soares. Average insert size 2-3kb.  
Note: this is a NIH\_MGC Library."

ORIGIN  
Query Match 33.8%; Score 829.8; DB 5; Length 916;  
Best Local Similarity 98.8%; Pred. No. 2.1e-165;  
Matches 856; Conservative 0; Mismatches 8; Indels 2; Gaps 2;  
Qy 610 GACCAAGCTCGTCGGGGAGGAGCCCTTTCCTTGGTGATTTCTGGGTCTGGCCGT 669  
Db 6 GCGCAAGCTCGTCGGGGAGGAGCCCTTTCCTTGGTGATTTCTGGGTCTGGCCGT 65  
Qy 670 GGGCTTCAGCTAGTCCAGGATGTTATCGCTATTGAATATATTGCTCCTCAATGAATAG 729  
Db 66 GGGCTTCAGCTAGTCCAGGATGTTATCGCTATTGAATATATTGCTCCTCAATGAATAG 125  
Qy 730 GACCAAGCTCATGCTTTCTGAGCTTTCGGCTCCTCGTCCGAATGAGACTTTGCTCT 789  
Db 126 GACCAAGCTCAATGCTTTCTGAGCTTTCGGCTCCTCGTCCGAATGAGACTTTGCTCT 185  
Qy 790 CTTGCTCAGCTACGCTCTCTTCTTGATGGCGCTGACCTTCTCTCATGCTCTCTTACCTT 849  
Db 186 CTTGCTCAGCTACGCTCTCTTCTTGATGGCGCTGACCTTCTCTCATGCTCTCTTACCTT 245  
Qy 850 CTGTGTTCTCTTCACGGGCTGGAAGAGACATCGGGCCCAATCTACCTCAAGATGCTCT 909  
Db 246 CTGTGTTCTCTTCACGGGCTGGAAGAGACATCGGGCCCAATCTACCTCAAGATGCTCT 305  
Qy 910 CTCGATGGCATCTGGTGGCTGGATCACCCTGCTCATGCTTCTCTGCTTTCAGCCGAG 969  
Db 306 CTCGATGGCATCTGGTGGCTGGATCACCCTGCTCATGCTTCTCTGCTTTCAGCCGAG 365  
Qy 970 GTGGATGACACCATCTCAGCTCCGCTTGGCTGCAATGGCTGGGTGTCTCTGTGGC 1029  
Db 366 GTGGATGACACCATCTCAGCTCCGCTTGGCTGCAATGGCTGGGTGTCTCTGTGGC 425  
Qy 1030 TTATGTTAGTCCGAGTTTGGCTGCTCAAAAGCAACGAAACCCCATGATATCTGT 1089  
Db 426 TTATGTTAGTCCGAGTTTGGCTGCTCAAAAGCAACGAAACCCCATGATATCTGT 485  
Qy 1090 TGAGGATGCTTCTGTAAACCTCAACTCGTGAAGAGAGCTATGTTGTGGAGAACAGGC 1149

[illegible]

RESULT 15	AL546138	839 bp	mRNA	linear	EST 25-MAR-2004
LOCUS	AL546138				
DEFINITION	AL546138 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens CDNA clone CSODI025YF10 5-PRIME, mRNA sequence.				
ACCESSION	AL546138				
VERSION	.AL546138.3	GI:45746606			
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1. (bases 1 to 839)				
AUTHORS	Li, W.B., Gruber, C., Jesse, J. and Polayes, D.				
TITLE	Full-length cDNA libraries and normalization				
JOURNAL	Unpublished (2001)				
COMMENT	On Feb 15, 2001 this sequence version replaced gi:31267972. Contact: Genoscope Genoscope - Centre National de Sequencage 2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime ends enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 6601.f For more information about this cluster, see <a href="http://www.genoscope.cns.fr/cdna?s=CSODI025DC050P1&amp;c=6601.f">http://www.genoscope.cns.fr/cdna?s=CSODI025DC050P1&amp;c=6601.f</a> .				

**FEATURES**  
**source**

## ORIGIN

### Query Match

33.4%; Score 820; DB 1; Length 839;

Best Local Similarity		99.2%	Pred. No. 2.5e-163;						
Matches	823;	Conservative	1;	Mismatches	6;	Indels	0;	Gaps	0;
Qy	160	GTCTCCCCCAGCACTGAGGAGCTCGCTGCTGCCCTCTTTCGCGCGGGGAAGCAGCACCAA	219						
Db	1	GTCTCCCCCAGCACTGAGGAGCTCGCTGCTGCCCTCTTTCGCGCGGGGAAGCAGCACCAA	60						
Qy	220	GTTTCAGGGCCAAACGCTTTGGCACTAGGGTCCAGAAATGGGTACAAAGTCCCTCATGGTTG	279						
Db	61	GTTTCAGGGCCAAACGCTTTGGCACTAGGGTCCAGAAATGGGTACAAAGTCCCTCATGGTTG	120						
Qy	280	CCGCAATGGCTCAAAATCCAAAGTACTACAGACTTTGTGATAAGGCTGAAGCTTGGGGCAT	339						
Db	121	CCGCAATGGCTCAAAATCCAAAGTACTACAGACTTTGTGATAAGGCTGAAGCTTGGGGCAT	180						
Qy	340	CGTCTTAGAAAACGGTGGCCACACGCGGGGTGTGACCTCGGTGGCCCTTCATGCTCACTCT	399						
Db	181	CGTCTTAGAAAACGGTGGCCACACGCTGGGTGTGTGACCTCAGTGGCCCTTCATGCTCACTCT	240						
Qy	400	CCCGATCTCTGCTCTGCAAGGTGCAGGACTCCAAACGCGGAAAATGCTGCCTACTCAGTT	459						
Db	241	CCCGATCTCTGCTCTGCAAGGTGCAGGACTCCAAACGCGGAAAATGCTGCCTACTCAGTT	300						
Qy	460	TCTCTTCTCTCTGGGTGTGTGGGCACTTTGGCCCTTCAGCTTCGCCTTCATCATCGGACT	519						
Db	301	TCTCTTCTCTCTGGGTGTGTGGGCACTTTGGCCCTTCAGCTTCGCCTTCATCATCGGACT	360						
Qy	520	GGACGGGAGCACAGGGCCACACAGCTTCTTCTCTTTTGGGATCCTCTTTTCCATCTGCTT	579						
Db	361	GGACGGGAGCACAGGGCCACACAGCTTCTTCTCTTTTGGGATCCTCTTTTCCATCTGCTT	420						
Qy	580	CTCTCGCTCTGGCTCATGCTGTGAGTGTGACCAAGTCTGTCGGGGAGGAAGGCCCT	639						
Db	421	CTCTCGCTCTGGCTCATGCTGTGAGTGTGACCAAGTCTGTCGGGGAGGAAGGCCCT	480						
Qy	640	TTCCCTGTTGGTGATTTCTGGGTCTGGCGTGGGCTTCAGCCTAGTCCAGGATCTTATCGC	699						
Db	481	TTCCCTGTTGGTGATTTCTGGGTCTGGCGTGGGCTTCAGCCTAGTCCAGGATCTTATCGC	540						
Qy	700	TATTGAATATATTGCTCTGACCATGAATAGGACCAACAGTCAATGTCTTTTCTCAGGCTTTC	759						
Db	541	TATTGAATATATTGCTCTGACCATGAATAGGACCAACAGTCAATGTCTTTCTCAGGCTTTC	600						
Qy	760	CGTCTCTCGTGCAGATGAAGAATTGTCCTCTGCTCACTAGTCCCTCTTCTTGATGGC	819						
Db	601	CGTCTCTCGTGCAGATGAAGAATTGTCCTCTGCTCACTAGTCCCTCTTCTTGATGGC	660						
Qy	820	GCTGACCTTCCCTCATGCTCTCTTACCCTTCTGTGGTTCCTTACAGGGCTGGAAGAGACA	879						
Db	661	GCTGACCTTCCCTCATGCTCTCTTACCCTTCTGTGGTTCCTTACAGGGCTGGAAGAGACA	720						
Qy	880	TGGGGGCCACATCACTACTCACGATGCTCTCTCCATTGGCCATCTGGGTGCCCTGGATCAC	939						
Db	721	TGGGGGCCACATCACTACTCAOGATGCTCTCTCCATTGGATCTGGGTGCCCTGGATCAC	780						
Qy	940	CTTGCTCATGCTTCTTGACCTTTGACCGCAGGTGGGATGACCAATCCTCA	989						
Db	781	CTTGCTCATGCTTCTTGACCTTTGACCGCAGTGGGATGACCAATCCTCA	830						

Search completed: June 10, 2006, 18:50:04  
Job time : 11520 secs

**THIS PAGE IS BLANK**





QY	181	CTGCGCTGCTGCGCTCTTTGCGCGGGAAGACAGACAAAGTTCAACGGCCAAACGCTTGGC	240
Db	181	CTGCGCTGCTGCGCTCTTTGCGCGGGAAGACAGACAAAGTTCAACGGCCAAACGCTTGGC	240
QY	241	ACTAGGTTCCAGAAATGGCTACACAGTCCCTGATGTTGCGCAATGGCTGAATCCAA	300
Db	241	ACTAGGTTCCAGAAATGGCTACACAGTCCCTGATGTTGCGCAATGGCTGAATCCAA	300
QY	301	GTACTACAGACTTTGTGATAAGGCTGAAGCTTGGGCGATCGTCTAGAAACGGTGGCCAC	360
Db	301	GTACTACAGACTTTGTGATAAGGCTGAAGCTTGGGCGATCGTCTAGAAACGGTGGCCAC	360
QY	361	AGCGGGGTTGTGACTCGGTGGCTTCAATGCTCACTCTCCGATCCTCGTCTGCAAGGT	420
Db	361	AGCGGGGTTGTGACTCGGTGGCTTCAATGCTCACTCTCCGATCCTCGTCTGCAAGGT	420
QY	421	GCAGGACTCCACAGCGGAATATGCTGCTACTCAGTTTCTCTTCTCTCTGCTGCTGTT	480
Db	421	GCAGGACTCCACAGCGGAATATGCTGCTACTCAGTTTCTCTTCTCTCTGCTGCTGTT	480
QY	481	GGGCATCTTTGGCTCACCTTCGCTTCATCATCGGACTGGACGGGACACAGGGCCAC	540
Db	481	GGGCATCTTTGGCTCACCTTCGCTTCATCATCGGACTGGACGGGACACAGGGCCAC	540
QY	541	ACGCTTCTCTCTTTGGGATCTCTTTTCCATCTGCTTCTGCTGCTGCTCATGC	600
Db	541	ACGCTTCTCTCTTTGGGATCTCTTTTCCATCTGCTTCTGCTGCTGCTCATGC	600
QY	601	TGTCAGTCTGACCAAGCTGTCGCGGGAGAGCCCTTCCCTGTTGGTGAATCTGGG	660
Db	601	TGTCAGTCTGACCAAGCTGTCGCGGGAGAGCCCTTCCCTGTTGGTGAATCTGGG	660
QY	661	TCGCGCGTGGGCTTCAGCCTAGTCCAGGATGTTATCGCTAATGAATATATTCCTGAC	720
Db	661	TCGCGCGTGGGCTTCAGCCTAGTCCAGGATGTTATCGCTAATGAATATATTCCTGAC	720
QY	721	CATGAATAGAACCAACGTCATGCTTTTCTGAGCTTTCGCTTCCTGCTGCGCAATGAAGA	780
Db	721	CATGAATAGAACCAACGTCATGCTTTTCTGAGCTTTCGCTTCCTGCTGCGCAATGAAGA	780
QY	781	CTTTGTCCTCTGCTCAGCTACCTGCTCTTCTTGTGATGGGCTGACCTTCTCATGTCCTC	840
Db	781	CTTTGTCCTCTGCTCAGCTACCTGCTCTTCTTGTGATGGGCTGACCTTCTCATGTCCTC	840
QY	841	CTTCACCTTCTGCTGCTTCTCAAGGCTGGAAGAGACATGCGGCGCCACATCTACCTCAC	900
Db	841	CTTCACCTTCTGCTGCTTCTCAAGGCTGGAAGAGACATGCGGCGCCACATCTACCTCAC	900
QY	901	GATGCTCCTCTCCATTTGGGCTGGGCTGGATCAGCTGCTCATGCTTCTGACTT	960
Db	901	GATGCTCCTCTCCATTTGGGCTGGGCTGGATCAGCTGCTCATGCTTCTGACTT	960
QY	961	TGACCGCAGTGGGATGACACCATCTCAGCTCCGCTTGGCTGCGCAATGGCTGGGTT	1020
Db	961	TGACCGCAGTGGGATGACACCATCTCAGCTCCGCTTGGCTGCGCAATGGCTGGGTT	1020
QY	1021	CCTGTTGGCTTATGTTAGTCCCGAGTTTGGCTGCTCACAAGAACCAACCCCATGGA	1080
Db	1021	CCTGTTGGCTTATGTTAGTCCCGAGTTTGGCTGCTCACAAGAACCAACCCCATGGA	1080
QY	1081	TTATCCTGTTGAGGATGCTTTCTGTAACCTCACTCAAGCTTGGTGAAGAGCTATGGTGA	1140
Db	1081	TTATCCTGTTGAGGATGCTTTCTGTAACCTCACTCAAGCTTGGTGAAGAGCTATGGTGA	1140
QY	1141	GAACAGAGCCTACTCTCAAGAGGAATCACTCAAGCTTGGTGAAGAGAGAGGACAGCT	1200
Db	1141	GAACAGAGCCTACTCTCAAGAGGAATCACTCAAGCTTGGTGAAGAGAGAGGACAGCT	1200
QY	1201	CTATGCCCCCTATTTCACACATTTTTCAGCTGCGAGAACCCAGCTCCTCCCAAGAAATTC	1260
Db	1201	CTATGCCCCCTATTTCACACATTTTTCAGCTGCGAGAACCCAGCTCCTCCCAAGAAATTC	1260

QY	1261	CATCCACCGGGCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAGAAAGAGG	1320
Db	1261	CATCCACCGGGCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAGAAAGAGG	1320
QY	1321	CAGCTAACTCTGTCTGAAGAGTGGGACAAATGACCGGGCGGACAGATCTACGGGAGC	1380
Db	1321	CAGCTAACTCTGTCTGAAGAGTGGGACAAATGACCGGGCGGACAGATCTACGGGAGC	1380
QY	1381	TCAAAGGAGTGGGGGAAATCTTGAGTCTTCTGAGAAAACCTGTACAAGACACTACGGGA	1440
Db	1381	TCAAAGGAGTGGGGGAAATCTTGAGTCTTCTGAGAAAACCTGTACAAGACACTACGGGA	1440
QY	1441	ACAGTTTGCTCCCTCCAGCTCAACCAAACTTCTTCCATGCTGGGCTGATGTGGCT	1500
Db	1441	ACAGTTTGCTCCCTCCAGCTCAACCAAACTTCTTCCATGCTGGGCTGATGTGGCT	1500
QY	1501	AGTAAGACTCCAGTTCTTAGAGCGCTGTAGTATTTTTTTTTTTTTTTTGTCTCATCTTTGG	1560
Db	1501	AGTAAGACTCCAGTTCTTAGAGCGCTGTAGTATTTTTTTTTTTTTTTTGTCTCATCTTTGG	1560
QY	1561	ATACTTTCTTTTAAAGTGGGAGTCTAGGCAAACTCAAGTTTAGACCTTACTCTTTTGT	1620
Db	1561	ATACTTTCTTTTAAAGTGGGAGTCTAGGCAAACTCAAGTTTAGACCTTACTCTTTTGT	1620
QY	1621	GTTTTTTGAACAGGATCTGTCTGTCAACCGGCTTGGTGCAGTGGTGGCATCACAG	1680
Db	1621	GTTTTTTGAACAGGATCTGTCTGTCAACCGGCTTGGTGCAGTGGTGGCATCACAG	1680
QY	1681	CCAGTGCAGCTCGACCACTGTGTCTCAAGCAATCTCCCATCTCCATCTCCCAAGTG	1740
Db	1681	CCAGTGCAGCTCGACCACTGTGTCTCAAGCAATCTCCCATCTCCATCTCCCAAGTG	1740
QY	1741	CTGGGATGACAGCGTGAAGCCACAGCTCCAGCTAGGCGCTTAAATCTTGTCTTATTT	1800
Db	1741	CTGGGATGACAGCGTGAAGCCACAGCTCCAGCTAGGCGCTTAAATCTTGTCTTATTT	1800
QY	1801	CCATGACTTAAAGTCTGGTCTAGCTCAGCTCAGCTGGCTCAGAGCTCTAGGGCGCTG	1860
Db	1801	CCATGACTTAAAGTCTGGTCTAGCTCAGCTCAGCTGGCTCAGAGCTCTAGGGCGCTG	1860
QY	1861	CTCTCTTAACTCAGTGGGTTTGTGAGCTCTGTGGGCCAGAGCAGACTCTCATATCT	1920
Db	1861	CTCTCTTAACTCAGTGGGTTTGTGAGCTCTGTGGGCCAGAGCAGACTCTCATATCT	1920
QY	1921	GAGCAAAAATAGCAAAAGCTCTCTCAGCCCACTGCGCTGAATCTACACTGGAAGCCAAC	1980
Db	1921	GAGCAAAAATAGCAAAAGCTCTCTCAGCCCACTGCGCTGAATCTACACTGGAAGCCAAC	1980
QY	1981	TTGCTGGCACCCCGCTCCCAACCTTCTTGCTGGTAGGAGAGCTAAAGATCACCC	2040
Db	1981	TTGCTGGCACCCCGCTCCCAACCTTCTTGCTGGTAGGAGAGCTAAAGATCACCC	2040
QY	2041	TAAATTTACTCATCTCTTAGTGTCTCACAATGGGCTCTCAGAGCTCTCCAGCACCA	2100
Db	2041	TAAATTTACTCATCTCTTAGTGTCTCACAATGGGCTCTCAGAGCTCTCCAGCACCA	2100
QY	2101	ATTACAGGTACCCCTCTCTTCTTGCACTGCCCCAACTTGTCTCAATTCGAGATC	2160
Db	2101	ATTACAGGTACCCCTCTCTTCTTGCACTGCCCCAACTTGTCTCAATTCGAGATC	2160
QY	2161	TAATCTCCCTTACGCTCTGCGAGGAATTTCTTTTTCAGACCTCAGTAGCACAAGCCGGT	2220
Db	2161	TAATCTCCCTTACGCTCTGCGAGGAATTTCTTTTTCAGACCTCAGTAGCACAAGCCGGT	2220
QY	2221	CTCTTTGTGAGGAAATTTGTAGATCAATCTCACTTCAAAATCTCTGGGCTGATCTCT	2280
Db	2221	CTCTTTGTGAGGAAATTTGTAGATCAATCTCACTTCAAAATCTCTGGGCTGATCTCT	2280
QY	2281	CTCATCTTGACCCCAACCTCTGTAATAGATTTACGGATTTACGGCTGCAATCTGTAA	2340
Db	2281	CTCATCTTGACCCCAACCTCTGTAATAGATTTACGGATTTACGGCTGCAATCTGTAA	2340
QY	2341	GTGGGATGCTCTCTTAAATGGAGAGTGTTCATTGTATAAAGTATTATTCACCTGAGTAT	2400

```
|||||
Db 2341 GTGGCGATGGTCTCTTAATGGAGAGTGTTCATTGTATATAAGTATTACCTGAGTAT 2400
Qy 2401 GCAATAAAGATGGTGGCCACTCTTTCATGGTGGTGGCGAGCAAAAAA 2456
Db 2401 GCAATAAAGATGGTGGCCACTCTTTCATGGTGGTGGCGAGCAAAAAA 2456

RESULT 2
DD210040 2456 bp DNA linear PAT 19-JAN-2006
LOCUS Methods of Diagnosis of Cancer, Compositions and Methods of
DEFINITION Screening for Modulators of Cancer.
ACCESSION DD210040
VERSION DD210040.1 GI-85654022
KEYWORDS JP 2005518782-A/34.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 2456)
AUTHORS Zlotnik,A., Mack,D.H., Agiz,N., Gish,K.C., Hebeji,P.A., Wilson,K.E.
and Afar,D.
TITLE Methods of Diagnosis of Cancer, Compositions and Methods of
Screening for Modulators of Cancer
JOURNAL Patent: JP 2005518782-A 34 30-JUN-2005;
PROTEIN DESIGN LABS INC
COMMENT OS Homo sapiens
PN JP 2005518782-A/34
PD 30-JUN-2005
PF 17-SEP-2002 JP 2003529912
PR 12-APR-2002 US 60/372246, 08-FEB-2002 US 60/355257, PR
08-FEB-2002 US 60/355145, 13-NOV-2001 US 60/350666, PR
20-SEP-2001 US 60/323887, 17-SEP-2001 US 60/323469 PI
albert
PI zlotnik,david h mack,natasha agiz,kurt c gish,peter a pi
hebeji,
CC keith e wilson,daniel afar
FH Location/Qualifiers.
source 1. .2456
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

Query Match 100.0%; Score 2456; DB 2; Length 2456;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ORIGIN
Qy 1 ATACAGCATGAGTGGCGTGGAACTGGAATAGCGGTCTCTCCCTCGACCTCCCC 60
Db 1 ATACAGCATGAGTGGCGTGGAACTGGAATAGCGGTCTCTCCCTCGACCTCCCC 60
Qy 61 TCCTTGTCCTCTGCTCACCCCTCGCTGTTTCCCTCCCTCGGCGAGGCGGCTTTATA 120
Db 61 TCCTTGTCCTCTGCTCACCCCTCGCTGTTTCCCTCCCTCGGCGAGGCGGCTTTATA 120
Qy 121 ACAACTGCTCAGAGTGGAGGGGGATAGCTGTCCTCAAGGTCTCCCGACCTGAGGAG 180
Db 121 ACAACTGCTCAGAGTGGAGGGGGATAGCTGTCCTCAAGGTCTCCCGACCTGAGGAG 180
Qy 181 CTGCGCTGTCCTCTCTTGGCGGGGAGAGCAGACCAAGTTACGCGCCACGCTTGGC 240
Db 181 CTGCGCTGTCCTCTCTTGGCGGGGAGAGCAGACCAAGTTACGCGCCACGCTTGGC 240
Qy 241 ACTAGGGTCCAGAAATGCTCAACAGTCCCTGATGTTGCGGCAATGGCTGAAATCCAA 300
Db 241 ACTAGGGTCCAGAAATGCTCAACAGTCCCTGATGTTGCGGCAATGGCTGAAATCCAA 300
Qy 301 GTACTACAGACTTTGTGATAGGCTGAAGCTTGGGGCATCGTCTAGAAACGGTGGCCAC 360
Db 301 GTACTACAGACTTTGTGATAGGCTGAAGCTTGGGGCATCGTCTAGAAACGGTGGCCAC 360
```

```
Qy 361 AGCCGGGTTGTGACCTCGTGGCCTTTCATGCTCACTCTCCCGATCCTCTCTGCAAGGT 420
Db 361 AGCCGGGTTGTGACCTCGTGGCCTTTCATGCTCACTCTCCCGATCCTCTCTGCAAGGT 420
Qy 421 GCAGGACTCCAAACAGGCGAAAAATGCTGCTACTCAGTCTTCTCTCTCTCTGCTGTT 480
Db 421 GCAGGACTCCAAACAGGCGAAAAATGCTGCTACTCAGTCTTCTCTCTCTCTGCTGTT 480
Qy 481 GGGCATCTTTGGGCTCACCTTCGCTTCATCATCGGACTCGGACGGAGGACACAGGCCAC 540
Db 481 GGGCATCTTTGGGCTCACCTTCGCTTCATCATCGGACTCGGACGGAGGACACAGGCCAC 540
Qy 541 ACGCTTCTCTCTCTTTGGGATCCTCTTTTCATCTGCTTCTCTCTCTCTCTCTCTCT 600
Db 541 ACGCTTCTCTCTCTTTGGGATCCTCTTTTCATCTGCTTCTCTCTCTCTCTCTCTCT 600
Qy 601 TGTCACTGACCAAGCTCGTCCGGGGAGGAGCCCTTTCCCTCTGTTGTTGTTCTGGG 660
Db 601 TGTCACTGACCAAGCTCGTCCGGGGAGGAGCCCTTTCCCTCTGTTGTTGTTCTGGG 660
Qy 661 TCTGGCGTGGGCTTCAGCCTAGTCAGGATGTTATCGCTATTGAAATATATTGTTCTGAC 720
Db 661 TCTGGCGTGGGCTTCAGCCTAGTCAGGATGTTATCGCTATTGAAATATATTGTTCTGAC 720
Qy 721 CATGAATAGGACCAAGCTCAATGCTCTTTTCTGAGCTTTCGCTCTCGTCCGAATGAAGA 780
Db 721 CATGAATAGGACCAAGCTCAATGCTCTTTTCTGAGCTTTCGCTCTCGTCCGAATGAAGA 780
Qy 781 CTTTGTCTCTCTCTCTCACCTACCTCTCTTTTGTGAGGCGCTGACCTTCTCATGTCTCTC 840
Db 781 CTTTGTCTCTCTCTCTCACCTACCTCTCTTTTGTGAGGCGCTGACCTTCTCATGTCTCTC 840
Qy 841 CTTCACTCTCTGTTGGTCTCTTACGGGCTGGAAAGAGACATGAGGGGCCACATCTACCTCAC 900
Db 841 CTTCACTCTCTGTTGGTCTCTTACGGGCTGGAAAGAGACATGAGGGGCCACATCTACCTCAC 900
Qy 901 GATGCTCTCTCTCCATTCGCTGAGTGGGCTGGATCAGCTCAGCTCTCATGCTCTCTGACTT 960
Db 901 GATGCTCTCTCTCCATTCGCTGAGTGGGCTGGATCAGCTCAGCTCTCATGCTCTCTGACTT 960
Qy 961 TGACCGAGTGGGATGACACCATCTCAGCTCCGCTTGGCTGCGCAATGGCTGGTGT 1020
Db 961 TGACCGAGTGGGATGACACCATCTCAGCTCCGCTTGGCTGCGCAATGGCTGGTGT 1020
Qy 1021 CTTTGGCTTATGTTAGTCCCGAGTTTGGCTGCTCAGCTCCGCTTGGCTGCGCAATGGCTGG 1080
Db 1021 CTTTGGCTTATGTTAGTCCCGAGTTTGGCTGCTCAGCTCCGCTTGGCTGCGCAATGGCTGG 1080
Qy 1081 TTATCCTGTTGAGGATGCTTTCTGTAACCTCAACTCGTGAAGAGAGCTATGTTGTTGA 1140
Db 1081 TTATCCTGTTGAGGATGCTTTCTGTAACCTCAACTCGTGAAGAGAGCTATGTTGTTGA 1140
Qy 1141 GAACAGAGCTTCTCTCAAGAGGAAATCACTCAAGGTTTGAAGAGACAGGGACAGCT 1200
Db 1141 GAACAGAGCTTCTCTCAAGAGGAAATCACTCAAGGTTTGAAGAGACAGGGACAGCT 1200
Qy 1201 CTATGCCCCCTTATTCACACATTTTTCAGCTGAGAACCCAGCTTCCCCAAAGAAATTTCTC 1260
Db 1201 CTATGCCCCCTTATTCACACATTTTTCAGCTGAGAACCCAGCTTCCCCAAAGAAATTTCTC 1260
Qy 1261 CATCCCAAGGCGGCGCTTGGCGGAGCCCTTACAAAGACTATGAAGTAAAGAAAGAGGG 1320
Db 1261 CATCCCAAGGCGGCGCTTGGCGGAGCCCTTACAAAGACTATGAAGTAAAGAAAGAGGG 1320
Qy 1321 CAGCTAACTCTGCTTGAAGAGTGGGACAAAATGACGCGGGCGGCGAGATCTAGCGGAGC 1380
Db 1321 CAGCTAACTCTGCTTGAAGAGTGGGACAAAATGACGCGGGCGGCGAGATCTAGCGGAGC 1380
Qy 1381 TCAAGGGATGTGGGCGAAATCTTTAGTCTTTCTGAGAAAACCTGTACAAGACACTACGGGA 1440
Db 1381 TCAAGGGATGTGGGCGAAATCTTTAGTCTTTCTGAGAAAACCTGTACAAGACACTACGGGA 1440
```

QY 1441 ACAGTTTGCTCCCTCCAGCCTCAACCACAAATCTTCCATGCTGGGCTGATGTGGCT 1500  
DB 1441 ACAGTTTGCTCCCTCCAGCCTCAACCACAAATCTTCCATGCTGGGCTGATGTGGCT 1500  
QY 1501 AGTAAGACTCCAGTCTCTAGAGCGCTGTAGTATTTTTTTTTTTTTTTTGTCTCATCTCTTGG 1560  
DB 1501 AGTAAGACTCCAGTCTCTAGAGCGCTGTAGTATTTTTTTTTTTTTTTTGTCTCATCTCTTGG 1560  
QY 1561 ATACTTTCTTTTAAGTGGAGTCTCAGGCAACTCAAGTTTAGACCCCTACTCTTTTGT 1620  
DB 1561 ATACTTTCTTTTAAGTGGAGTCTCAGGCAACTCAAGTTTAGACCCCTACTCTTTTGT 1620  
QY 1621 GTTTTTTGAACAGGATCTGCTCTGTACCCAGGCTTGAGTCAGTGGTGGATCACAG 1680  
DB 1621 GTTTTTTGAACAGGATCTGCTCTGTACCCAGGCTTGAGTCAGTGGTGGATCACAG 1680  
QY 1681 CCAGTGCAGCCTCGACCACTGTCTGCTCAAGCAATCTCCCATCTCCATCTCCCAAGTG 1740  
DB 1681 CCAGTGCAGCCTCGACCACTGTCTGCTCAAGCAATCTCCCATCTCCATCTCCCAAGTG 1740  
QY 1741 CTGGGATGACAGCGTGTGAGCCAGCTCCAGCCTAGGCGCTTAATCTTGTCTTATTTT 1800  
DB 1741 CTGGGATGACAGCGTGTGAGCCAGCTCCAGCCTAGGCGCTTAATCTTGTCTTATTTT 1800  
QY 1801 CCATGACTAAAGTCTGTGATCTGAGCTCAGCTGGCTCAGACAGCTCTAGGGCCCTG 1860  
DB 1801 CCATGACTAAAGTCTGTGATCTGAGCTCAGCTGGCTCAGACAGCTCTAGGGCCCTG 1860  
QY 1861 CTCTCTTAACCTACAGTGGGTTTGTGAGGCTCTGTGGCCAGAGCAGCTCATATCT 1920  
DB 1861 CTCTCTTAACCTACAGTGGGTTTGTGAGGCTCTGTGGCCAGAGCAGCTCATATCT 1920  
QY 1921 GAGCAAAAATAGCAAAAGCTCTCTCAGCCCACTGCGCTGAATCTACCTGGAAGCAAC 1980  
DB 1921 GAGCAAAAATAGCAAAAGCTCTCTCAGCCCACTGCGCTGAATCTACCTGGAAGCAAC 1980  
QY 1981 TTGCTGGCACCCTGCTCCCACTCTTCTGCTGGGTAGGAGGCTTAAGATCACCC 2040  
DB 1981 TTGCTGGCACCCTGCTCCCACTCTTCTGCTGGGTAGGAGGCTTAAGATCACCC 2040  
QY 2041 TAAATTTACTCATCTCTTAGTGTGCTCCTCACAATGGGCTCAGCAGCTCCCAAGCA 2100  
DB 2041 TAAATTTACTCATCTCTTAGTGTGCTCCTCACAATGGGCTCAGCAGCTCCCAAGCA 2100  
QY 2101 ATTACAGGTCACCCCTCTCTTTGAGCTGTGCTCCCAAACTTGTCTCAATTCGGAGATC 2160  
DB 2101 ATTACAGGTCACCCCTCTCTTTGAGCTGTGCTCCCAAACTTGTCTCAATTCGGAGATC 2160  
QY 2161 TAATCTCCCTTACGCTCTGCGAGGAATCTTTTCAGACCTCAGTACAGCAAGCCGGTGG 2220  
DB 2161 TAATCTCCCTTACGCTCTGCGAGGAATCTTTTCAGACCTCAGTACAGCAAGCCGGTGG 2220  
QY 2221 CTCTTTGTGAGGAGAAATTTGTAGATCATCTCACTTCAAAATCTGGGCTGATCTTCT 2280  
DB 2221 CTCTTTGTGAGGAGAAATTTGTAGATCATCTCACTTCAAAATCTGGGCTGATCTTCT 2280  
QY 2281 CTCTTTGTGAGGAGAAATTTGTAGATCATCTCACTTCAAAATCTGGGCTGATCTTCT 2340  
DB 2281 CTCTTTGTGAGGAGAAATTTGTAGATCATCTCACTTCAAAATCTGGGCTGATCTTCT 2340  
QY 2341 GTGGGATGCTCTTAATGGAGGAGTGTCTTATGATTAATTAAGTTTATTCACCTGAGTAT 2400  
DB 2341 GTGGGATGCTCTTAATGGAGGAGTGTCTTATGATTAATTAAGTTTATTCACCTGAGTAT 2400  
QY 2401 GCAATAAAGATGTGGTGGCCACTCTTTTCATGTTGGTGGCAGCAAAAAA 2456  
DB 2401 GCAATAAAGATGTGGTGGCCACTCTTTTCATGTTGGTGGCAGCAAAAAA 2456

RESULT 3  
AX549168  
LOCUS  
DEFINITION Sequence 453 from Patent WO02061087.  
AX549168 2456 bp DNA linear PAT 26-NOV-2002

ACCESSION AX549168  
VERSION AX549168.1 GI:25813894  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
REFERENCE 1  
AUTHORS Burner,G.C., Roush,C.L. and Brown,J.P.  
TITLE Antigenic peptides, such as for G protein-coupled receptors (GPCRs), antibodies thereto, and systems for identifying such antigenic peptides  
JOURNAL Patent: WO 02061087-A 453 08-AUG-2002;  
Lifespan Biosciences, Inc. (US)  
FEATURES  
source  
1..2456  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Query Match 100.0%; Score 2456; DB 2; Length 2456;  
Best Local Similarity 100.0%; Pred No. 0;  
Matches 2456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATAAACAGCATGAAGTGCCTGGAACCTGGAATAGGCGTGTCTCTCCCTGACCCCTCCCC 60  
DB 1 ATAAACAGCATGAAGTGCCTGGAACCTGGAATAGGCGTGTCTCTCCCTGACCCCTCCCC 60  
QY 61 TCTTTGTCCCTCTGCTACCCCTCGCTCCCTCCCTCCCTCCGCGAGGCGCCGCTTTATA 120  
DB 61 TCTTTGTCCCTCTGCTACCCCTCGCTCCCTCCCTCCCTCCGCGAGGCGCCGCTTTATA 120  
QY 121 ACAACTGCTCAGAGTGCAGGCGGGGATAGTGTCCAAGTCTCCCGCCAGCACTGAGGAG 180  
DB 121 ACAACTGCTCAGAGTGCAGGCGGGGATAGTGTCCAAGTCTCCCGCCAGCACTGAGGAG 180  
QY 181 CTCGCCCTGTGCCCTTTTGGCGGGAAGCAGCACAAGTTTCAGCGCCAAACGCTTGGC 240  
DB 181 CTCGCCCTGTGCCCTTTTGGCGGGAAGCAGCACAAGTTTCAGCGCCAAACGCTTGGC 240  
QY 241 ACTAGGGTCCAGAAATGGCTACAAACAGTCCCTGATGTTTCCGCAATGGCTGAAATCCAA 300  
DB 241 ACTAGGGTCCAGAAATGGCTACAAACAGTCCCTGATGTTTCCGCAATGGCTGAAATCCAA 300  
QY 301 GTACTACAGACTTTGTGATAAGCTTTGGGCAATCGTCTTAGAAAACGCTGGCCAC 360  
DB 301 GTACTACAGACTTTGTGATAAGCTTTGGGCAATCGTCTTAGAAAACGCTGGCCAC 360  
QY 361 AGCGGGGTGTGACCTCGGTGGCTTCACTGCTCACTCTCCGATCCTGCTGTGCAAGGT 420  
DB 361 AGCGGGGTGTGACCTCGGTGGCTTCACTGCTCACTCTCCGATCCTGCTGTGCAAGGT 420  
QY 421 GCAGGACTCCAAACAGCGGAAATGCTGCTACTCAGTTTCTCTCTCTGGGTGTGT 480  
DB 421 GCAGGACTCCAAACAGCGGAAATGCTGCTACTCAGTTTCTCTCTCTGGGTGTGT 480  
QY 481 GGCGATCTTTGGCCTCACCTTCGCTTCATCGGACTGGAGGGGAGCAGAGGGCCAC 540  
DB 481 GGCGATCTTTGGCCTCACCTTCGCTTCATCGGACTGGAGGGGAGCAGAGGGCCAC 540  
QY 541 AGCTTTCTCTCTTTGGGATCTCTTTTCCATCTGCTTCTCTGCTGCTGCTCATGC 600  
DB 541 AGCTTTCTCTCTTTGGGATCTCTTTTCCATCTGCTTCTCTGCTGCTGCTCATGC 600  
QY 601 TGTTCAGTCTGACCAAGCTGTCTCGGGGAGGAGGCGCTTTCCCTGTTGGTGTATTCTGGG 660  
DB 601 TGTTCAGTCTGACCAAGCTGTCTCGGGGAGGAGGCGCTTTCCCTGTTGGTGTATTCTGGG 660  
QY 661 TCTGGCGGTGGGCTTCAGCCTAGTCCAGGATGTTATCGGTATTTGAATATATTCTCTGAC 720  
DB 661 TCTGGCGGTGGGCTTCAGCCTAGTCCAGGATGTTATCGGTATTTGAATATATTCTCTGAC 720

QY	721	CATGAATAGGACCAACGTCATATGCTCTTTCTGAGCTTTCCGCTCTCTCGTCGCAATGAAGA	780
DB	721	CATGAATAGGACCAACGTCATATGCTCTTTCTGAGCTTTCCGCTCTCTCGTCGCAATGAAGA	780
QY	781	CTTTGTCTCTGCTGCTACCTACGTCCTCTCTTTGATGGCGTGACCTTCTCTCATGTCCCTC	840
DB	781	CTTTGTCTCTGCTGCTACCTACGTCCTCTCTTTGATGGCGTGACCTTCTCTCATGTCCCTC	840
QY	841	CTTCACCTCTCTGCTGCTACGCGCTGGAAGACATGCGGCGCCACATCTACCTCAC	900
DB	841	CTTCACCTCTCTGCTGCTACGCGCTGGAAGACATGCGGCGCCACATCTACCTCAC	900
QY	901	GATCCTCTCTCCATTCGCCATCTGGGTGGCTGGATCACCTCTCATGCTTCTCTGACTT	960
DB	901	GATCCTCTCTCCATTCGCCATCTGGGTGGCTGGATCACCTCTCATGCTTCTCTGACTT	960
QY	961	TGACCGGAGGTGGATGACACCAATCTCAGCTCCGCTTGGCTGCGCAATGGCTGGGTGTT	1020
DB	961	TGACCGGAGGTGGATGACACCAATCTCAGCTCCGCTTGGCTGCGCAATGGCTGGGTGTT	1020
QY	1021	CTGTGGCTTATGTTAGTCCGAGTTTGGCTGCTCAGAAAGCAACGAAACCCCATGGA	1080
DB	1021	CTGTGGCTTATGTTAGTCCGAGTTTGGCTGCTCAGAAAGCAACGAAACCCCATGGA	1080
QY	1081	TTATCTCTGTTGAGGATGCTTTCTGTAACCTCAACTCGTGAAGAGAGCTATGTTGGA	1140
DB	1081	TTATCTCTGTTGAGGATGCTTTCTGTAACCTCAACTCGTGAAGAGAGCTATGTTGGA	1140
QY	1141	GAACAGAGCTTACTCTCAAGAGGAAATCACTCAAGGTTTTGAAGAGACAGGGACACGCT	1200
DB	1141	GAACAGAGCTTACTCTCAAGAGGAAATCACTCAAGGTTTTGAAGAGACAGGGACACGCT	1200
QY	1201	CTATGCCCCCTATTTCCACATTTTCACTGAGAACAGCCAGCTCCCAAGAAATTTCTC	1260
DB	1201	CTATGCCCCCTATTTCCACATTTTCACTGAGAACAGCCAGCTCCCAAGAAATTTCTC	1260
QY	1261	CATCCACGGGCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAGAAAGGG	1320
DB	1261	CATCCACGGGCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAGAAAGGG	1320
QY	1321	CAGTAACTCTGTCTGAAGAGTGGGACAAATGAGCGCGGCGGCGAGATCTAGCGGAGC	1380
DB	1321	CAGTAACTCTGTCTGAAGAGTGGGACAAATGAGCGCGGCGGCGAGATCTAGCGGAGC	1380
QY	1381	TCAAAGGGATGTTGGCGAAATCTTGAAGTCTTCTGAGAAACTGTGAAGACATACCGGA	1440
DB	1381	TCAAAGGGATGTTGGCGAAATCTTGAAGTCTTCTGAGAAACTGTGAAGACATACCGGA	1440
QY	1441	ACAGTTTGCCTCCCTCCAGCCTCAACCAATTTCTTCCATGCTGGGCTGATGTGGCT	1500
DB	1441	ACAGTTTGCCTCCCTCCAGCCTCAACCAATTTCTTCCATGCTGGGCTGATGTGGCT	1500
QY	1501	AGTAAGACTCCAGTTCTTAGAGGCGCTGTAGTATTTTTTTTTTTTGTCTCATCCTTTGG	1560
DB	1501	AGTAAGACTCCAGTTCTTAGAGGCGCTGTAGTATTTTTTTTTTTTGTCTCATCCTTTGG	1560
QY	1561	ATACTCTTTTAAAGTGGAGTCTCAGGCAACTCAAGTTTGAACCCCTTACTCTTTTGT	1620
DB	1561	ATACTCTTTTAAAGTGGAGTCTCAGGCAACTCAAGTTTGAACCCCTTACTCTTTTGT	1620
QY	1621	GTTTTTTGAACAGGATCTTGTCTGTACCCAGGCTTGAAGTGAAGTGGGATCACAG	1680
DB	1621	GTTTTTTGAACAGGATCTTGTCTGTACCCAGGCTTGAAGTGAAGTGGGATCACAG	1680
QY	1681	CCCAAGTCCAGCTCGACCACTGTGCTCAAGCAATCTCCCATCTCCCATCTCCCAAGTG	1740
DB	1681	CCCAAGTCCAGCTCGACCACTGTGCTCAAGCAATCTCCCATCTCCCATCTCCCAAGTG	1740
QY	1741	CTGGGATGACAGGCGTGAAGCCACAGCTCCAGCCTTAGGCCCTTAATCTTCTGTATTTT	1800
DB	1741	CTGGGATGACAGGCGTGAAGCCACAGCTCCAGCCTTAGGCCCTTAATCTTCTGTATTTT	1800

Qy	1801	CCATGGACTAAAGGTC	TGGTCACTCTGAGCTCAGCTGGCTCACACAGCTCTAGGGGCTG	1860
Db	1801	CCATGGACTAAAGGTC	TGGTCACTCTGAGCTCAGCTGGCTCACACAGCTCTAGGGGCTG	1860
Qy	1861	CTCCTCTAACTCAAGTGGGTTTGTGAGGCTCTGTGGCCAGAGCAGACCTGCATATCT		1920
Db	1861	CTCCTCTAACTCAAGTGGGTTTGTGAGGCTCTGTGGCCAGAGCAGACCTGCATATCT		1920
Qy	1921	GAGCAAAATAGCAAAAGCCTCTCTCAGCCCACTGGCTGAATCTACACTGGAAGCCAAC		1980
Db	1921	GAGCAAAATAGCAAAAGCCTCTCTCAGCCCACTGGCTGAATCTACACTGGAAGCCAAC		1980
Qy	1981	TTGCTGGCACCCCGCTCCCAACCTCTTCTGCTGGTAGGAGAGCTAAAGATCACCC		2040
Db	1981	TTGCTGGCACCCCGCTCCCAACCTCTTCTGCTGGTAGGAGAGCTAAAGATCACCC		2040
Qy	2041	TAAATTTACTCATCTCTCTAGTGTGCTCAGATGCTCAGAGCTCCCGAGATC		2100
Db	2041	TAAATTTACTCATCTCTCTAGTGTGCTCAGATGCTCAGAGCTCCCGAGATC		2100
Qy	2101	ATTACAGGTCAACCCCTCTCTTCTTGCACTGTCCCAAACTTCTGCTCAATTCGAGATC		2160
Db	2101	ATTACAGGTCAACCCCTCTCTTCTTGCACTGTCCCAAACTTCTGCTCAATTCGAGATC		2160
Qy	2161	TAACTCCCTTACGCTCTGCCAGGAATCTTTTTCAGACCTCAGTACACAGCCCGGTTG		2220
Db	2161	TAACTCCCTTACGCTCTGCCAGGAATCTTTTTCAGACCTCAGTACACAGCCCGGTTG		2220
Qy	2221	CTCCTTGTGAGGAGAAATTTGTAGATCAATTTCTCACTTCAAATTCCTGGGCTGATCTTCT		2280
Db	2221	CTCCTTGTGAGGAGAAATTTGTAGATCAATTTCTCACTTCAAATTCCTGGGCTGATCTTCT		2280
Qy	2281	CTCATCTTGACCCCAACCTCTGTAATAGATTTACGGATTTACGGCTGCATCTGTGTA		2340
Db	2281	CTCATCTTGACCCCAACCTCTGTAATAGATTTACGGATTTACGGCTGCATCTGTGTA		2340
Qy	2341	GTGGGATGCTCTCTTAATGGAGAGTGTTCATTGTATAAAGTATTACACCTGAGTAT		2400
Db	2341	GTGGGATGCTCTCTTAATGGAGAGTGTTCATTGTATAAAGTATTACACCTGAGTAT		2400
Qy	2401	GCAATAAAGATGTGTGGCCACTCTTTTCATGTGTGGCAGCAAAAAA		2456
Db	2401	GCAATAAAGATGTGTGGCCACTCTTTTCATGTGTGGCAGCAAAAAA		2456
RESULT 4				
CO894732				
LOCUS	6730 bp DNA linear PAT 05-NOV-2004			
DEFINITION	Sequence 42 from Patent EP1471075.			
ACCESSION	CO894732			
VERSION	CO894732.1 GI:55467481			
KEYWORDS	Homo sapiens (human)			
SOURCE	Homo sapiens			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE	1			
AUTHORS	Rosenthal, A.D., Pillarsky, C., Dahl, E., Specht, T., Brummendorf, T., Lichtner, R., Staub, E., Roepcke, S. and Li, X.I.			
TITLE	Human nucleic acid sequences expressed in pancreatic carcinomas			
JOURNAL	Patent: EP 1471075-A 42 27-OCT-2004; Hinzmann, Bernd (DE); Rosenthal, Andre (DE); Pillarsky, Christian (DE); Dahl, Edgar (DE); Specht, Thomas (DE); Lichtner, Rosemarie (DE)			
FEATURES	Location/Qualifiers			
source	1..6730			
	/organism="Homo sapiens"			
	/mol_type="unassigned DNA"			
	/db_xref="taxon:9606"			
ORIGIN				

Best Local Similarity 100.0%; Pred. No. 0; Matches 2443; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	ATAACAGCATGAAGTCGCGTGAACCTGGAACTAGCGGTGTCCTCTCCCTCGACCCCTCCCC	60						
Db	1	ATAACAGCATGAAGTCGCGTGAACCTGGAACTAGCGGTGTCCTCTCCCTCGACCCCTCCCC	60						
QY	61	TCCTTGTCCCTCTGCTCAACCCCTCGCTGCTTCCCTCCCTCCGCGAGGCGCGCTTTATA	120						
Db	61	TCCTTGTCCCTCTGCTCAACCCCTCGCTGCTTCCCTCCCTCCGCGAGGCGCGCTTTATA	120						
QY	121	ACAACTGCTCAGAGTCGCGGGCGGATAGCTGTCGAAGTCTCCGCCAGCTAGAGGAG	180						
Db	121	ACAACTGCTCAGAGTCGCGGGCGGATAGCTGTCGAAGTCTCCGCCAGCTAGAGGAG	180						
QY	181	CTCGCCTGTGCTCCCTCTTGCGCGCGGAAGCAGCACAAGTTCAACGGCCAAACGCTTTGGC	240						
Db	181	CTCGCCTGTGCTCCCTCTTGCGCGCGGAAGCAGCACAAGTTCAACGGCCAAACGCTTTGGC	240						
QY	241	ACTAGGGTCCAGAAATGGCTACAACAGTCCCTGATGTTGCGCAATGGCTGAAATCCAA	300						
Db	241	ACTAGGGTCCAGAAATGGCTACAACAGTCCCTGATGTTGCGCAATGGCTGAAATCCAA	300						
QY	301	GTACTACAGACTTTGTGATAAGCTCAAGCTTGGGCATCGTCTAGAAACGGTGGCCAC	360						
Db	301	GTACTACAGACTTTGTGATAAGCTCAAGCTTGGGCATCGTCTAGAAACGGTGGCCAC	360						
QY	361	AGCGGGGTTGTGACCTCGGTGGCTTCATGCTCACTCTCCCGATCCTGCTGCAAGGT	420						
Db	361	AGCGGGGTTGTGACCTCGGTGGCTTCATGCTCACTCTCCCGATCCTGCTGCAAGGT	420						
QY	421	GCAGGACTCCAAACAGCGGAAAATGTGCTACTCAAGTTCCTCTCTGCGGTGTGT	480						
Db	421	GCAGGACTCCAAACAGCGGAAAATGTGCTACTCAAGTTCCTCTCTGCGGTGTGT	480						
QY	481	GGGCATCTTTGGCCTCACCTTCGCTTCACTCGGACTGGAGCGGAGCACAGGGCCAC	540						
Db	481	GGGCATCTTTGGCCTCACCTTCGCTTCACTCGGACTGGAGCGGAGCACAGGGCCAC	540						
QY	541	ACGCTTCTTCCCTCTTTGGGATCCTCTTTTCCATCTGCTTCTCTGCTGCTCATGC	600						
Db	541	ACGCTTCTTCCCTCTTTGGGATCCTCTTTTCCATCTGCTTCTCTGCTGCTCATGC	600						
QY	601	TGTCAGTCTGAACCAAGCTGTGCGGGGAGGAAGCCCTTTCCTGTTGGTGAATCTGGG	660						
Db	601	TGTCAGTCTGAACCAAGCTGTGCGGGGAGGAAGCCCTTTCCTGTTGGTGAATCTGGG	660						
QY	661	TCGCGCGTGGGCTTCAGCCTAGTCAGAGATGTTATCGCTATTGAATATATTGCTCTGAC	720						
Db	661	TCGCGCGTGGGCTTCAGCCTAGTCAGAGATGTTATCGCTATTGAATATATTGCTCTGAC	720						
QY	721	CATGAATAGGACCAACGTCATGCTCTTTTCTGAGCTTTCGGCTCTCGTCGCAATGAAGA	780						
Db	721	CATGAATAGGACCAACGTCATGCTCTTTTCTGAGCTTTCGGCTCTCGTCGCAATGAAGA	780						
QY	781	CTTTGCTCTGCTCAGCTACCTCTCTCTTTGATGGGCTGACCTTCTCATGTCTTC	840						
Db	781	CTTTGCTCTGCTCAGCTACCTCTCTCTTTGATGGGCTGACCTTCTCATGTCTTC	840						
QY	841	CTTTACCTTCTGTGGTTCTTCAAGGGCTGGAAGAGACATGGGGCCACATCTACCTCAC	900						
Db	841	CTTTACCTTCTGTGGTTCTTCAAGGGCTGGAAGAGACATGGGGCCACATCTACCTCAC	900						
QY	901	GATGCTCTCTCAATTTGCCATCTGGGTGGCTGGATCACCTGCTCATGCTTCTGACTT	960						
Db	901	GATGCTCTCTCAATTTGCCATCTGGGTGGCTGGATCACCTGCTCATGCTTCTGACTT	960						
QY	961	TGACCCAGGTGGGATGACACATCTCAGCTCCGCTTGGCTGGCCAAATGGCTGGGTGT	1020						
Db	961	TGACCCAGGTGGGATGACACATCTCAGCTCCGCTTGGCTGGCCAAATGGCTGGGTGT	1020						
QY	1021	CTGTTGGCTTATGTTAGTCCCGAGTTTGGCTGCTCACAAGCAACGAAACCCCATGGA	1080						

Db	1021	CTGTTGGCTTATGTTAGTCCCGAGTTTGGCTGCTCACAAGCAACGAAACCCCATGGA	1080
QY	1081	TTATCCTGTTGAGGATGCTTTCTGTAAACCTCAACTCGTGAAGAGAGCTATGGTGTGA	1140
Db	1081	TTATCCTGTTGAGGATGCTTTCTGTAAACCTCAACTCGTGAAGAGAGCTATGGTGTGA	1140
QY	1141	GAACAGAGCTACTCTCAAGAGGAAATCACTCAAGTCTTTGAAGACAGGGGACAGCT	1200
Db	1141	GAACAGAGCTACTCTCAAGAGGAAATCACTCAAGTCTTTGAAGACAGGGGACAGCT	1200
QY	1201	CTATGCCCCCTATTCACACATTTTCAGCTGCAAGAACAGCCTCCCCAAAAGGAATCTC	1260
Db	1201	CTATGCCCCCTATTCACACATTTTCAGCTGCAAGAACAGCCTCCCCAAAAGGAATCTC	1260
QY	1261	CATCCACAGGGCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAGAAAGAGG	1320
Db	1261	CATCCACAGGGCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAGAAAGAGG	1320
QY	1321	CAGCTAACTCTGCTGAAAGAGTGGGACAAATGTCAGCCGGCGGAGATCTAGCGGAGC	1380
Db	1321	CAGCTAACTCTGCTGAAAGAGTGGGACAAATGTCAGCCGGCGGAGATCTAGCGGAGC	1380
QY	1381	TCAAAGGGATGTGGCGGAAATCTTGAGTCTTCTGAGAAAACCTGACAAACACTACGGGA	1440
Db	1381	TCAAAGGGATGTGGCGGAAATCTTGAGTCTTCTGAGAAAACCTGACAAACACTACGGGA	1440
QY	1441	ACAGTTTGCTTCCCTCCAGCCTCAACCAAACTTCTTCATGCTGGGCTGATGGGCT	1500
Db	1441	ACAGTTTGCTTCCCTCCAGCCTCAACCAAACTTCTTCATGCTGGGCTGATGGGCT	1500
QY	1501	AGTAAGACTCAGTCTTTAGAGCGCTGAGTATTTTTTTTTTTTTTTTGTCTCATCTCTGG	1560
Db	1501	AGTAAGACTCAGTCTTTAGAGCGCTGAGTATTTTTTTTTTTTTTTTGTCTCATCTCTGG	1560
QY	1561	ATACTTCTTTTAAAGTGGGAGTCTCAGGCAACTCAAGTTTAGACCCCTTACTCTTTTGT	1620
Db	1561	ATACTTCTTTTAAAGTGGGAGTCTCAGGCAACTCAAGTTTAGACCCCTTACTCTTTTGT	1620
QY	1621	GTTTTTTGAAAACAGGATCTTGTCACCCAGGCTTGAGTGCAGTGTGGGATCACAG	1680
Db	1621	GTTTTTTGAAAACAGGATCTTGTCACCCAGGCTTGAGTGCAGTGTGGGATCACAG	1680
QY	1681	CCAGTGCAGCCTCGACACCTGCTCAAGCAATCCTCCCATCTCCCAAGTG	1740
Db	1681	CCAGTGCAGCCTCGACACCTGCTCAAGCAATCCTCCCATCTCCCAAGTG	1740
QY	1741	CTGGGATGACAGGCGTGAGCCACAGCTCCAGCCTTAGGCCCTTAATCTTGTGTTATTT	1800
Db	1741	CTGGGATGACAGGCGTGAGCCACAGCTCCAGCCTTAGGCCCTTAATCTTGTGTTATTT	1800
QY	1801	CCATGGAATAAAGGCTTGTCATCTGAGCTCAGCTGGGCTCACAGCTCTAGGGGCTG	1860
Db	1801	CCATGGAATAAAGGCTTGTCATCTGAGCTCAGCTGGGCTCACAGCTCTAGGGGCTG	1860
QY	1861	CTCCTCTACTCAGTGGGTTTTTGAGGCTCTGTGGCCAGAGCAGACCTGCATATCT	1920
Db	1861	CTCCTCTACTCAGTGGGTTTTTGAGGCTCTGTGGCCAGAGCAGACCTGCATATCT	1920
QY	1921	GAGCAAAAATAGCAAAAGCCTCTCTCAGCCCACTGGCCTGAATCTACACTGGAAGCAAC	1980
Db	1921	GAGCAAAAATAGCAAAAGCCTCTCTCAGCCCACTGGCCTGAATCTACACTGGAAGCAAC	1980
QY	1981	TTGCTGGACCCCGCTCCCAACCCCTTCTGCTGGGTAGGAGGCTTAAGATCACCC	2040
Db	1981	TTGCTGGACCCCGCTCCCAACCCCTTCTGCTGGGTAGGAGGCTTAAGATCACCC	2040
QY	2041	TAAATTTACTCATCTCTCTAGTGCCTCAATTTGGGCTCAGCAGCTCCCGACCA	2100
Db	2041	TAAATTTACTCATCTCTCTAGTGCCTCAATTTGGGCTCAGCAGCTCCCGACCA	2100
QY	2101	ATTACAGGTACCCCTCTCTTTGCACTGTCCCAAACTTGTCTCAATTCGAGATC	2160
Db	2101	ATTACAGGTACCCCTCTCTTTGCACTGTCCCAAACTTGTCTCAATTCGAGATC	2160

```

QY 2161 TAATCTCCCTACGCTCTGCGAGGAATCTTTTCAGACCTCACTAGCACAAAGCCCGGTTG 2220
Db 2161 TAATCTCCCTACGCTCTGCGAGGAATCTTTTCAGACCTCACTAGCACAAAGCCCGGTTG 2220
QY 2221 CTCTTGTTCAGGAGAAATTTGTAGATCATCTCACTTCAAAATTCCTGGGGCTGATCTTCT 2280
Db 2221 CTCTTGTTCAGGAGAAATTTGTAGATCATCTCACTTCAAAATTCCTGGGGCTGATCTTCT 2280
QY 2281 CTATCTTGACCCCAACCTCTGTAATAGATTTACCGCAATTTACGGCTCATCTGTAA 2340
Db 2281 CTATCTTGACCCCAACCTCTGTAATAGATTTACCGCAATTTACGGCTCATCTGTAA 2340
QY 2341 GTGGGATGCTCTCCTAAATGGAGAGTGTTCATTTGTATATAAGTTATTACCTGAGTAT 2400
Db 2341 GTGGGATGCTCTCCTAAATGGAGAGTGTTCATTTGTATATAAGTTATTACCTGAGTAT 2400
QY 2401 GCAATAAAGATGTGGTGGCCACTCTTTTCATGTGGTGGCAGCA 2443
Db 2401 GCAATAAAGATGTGGTGGCCACTCTTTTCATGTGGTGGCAGCA 2443

RESULT 5
BD156680
LOCUS
DEFINITION 2446 bp DNA linear PAT 17-JAN-2003
ACCESSION Primer for synthesizing full-length cDNA and use thereof.
VERSION BD156680.1 GI:27862438
KEYWORDS JP 2002191363-A/11523.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
TITLE Primer for synthesizing full-length cDNA and use thereof
JOURNAL Patent: JP 2002191363-A 11523 09-JUL-2002;
HELIX RESEARCH INSTITUTE
COMMENT OS Homo sapiens (human)
PN JP 2002191363-A/11523
PD 09-JUL-2002
PF 28-JUL-2000 JP 2000280990
PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU
SAITO,
PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,
PI KEIICHI NAGAI, TETSUJI OTSUKI
PC
C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC
10, C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key
Location/Qualifiers
FT CDS (254)..(1324).
source 1..2446
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 99.3%; Score 2439.8; DB 2; Length 2446;
Best Local Similarity 99.94; Pred: No. 0;
Matches 2441; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATACAGCATGAAGTCGCGTGAACCTGGAATAGCGGTCTCTCCCTCGACCCCTCCCC 60
Db 1 ATACAGCATGAAGTCGCGTGAACCTGGAATAGCGGTCTCTCCCTCGACCCCTCCCC 60
QY 61 TCCTTGTCCTCTGCTCACCCCTCGCTCGTTCCTCCCTCCGCGAGGGCGGCTTTATA 120
Db 61 TCCTTGTCCTCTGCTCACCCCTCGCTCGTTCCTCCCTCCGCGAGGGCGGCTTTATA 120

```

```

QY 121 ACAACTGCTCAGAGTCGAGGCGGATAGCTGTCCAAGGTCTCCCCAGCACTGAGGAG 180
Db 121 ACAACTGCTCAGAGTCGAGGCGGATAGCTGTCCAAGGTCTCCCCAGCACTGAGGAG 180
QY 181 CTGCGCTGCTGCCCTCTTTGGCGCGGGAAGCAGCACCAAGTTTTCACGGCCAAACGCTTTGGC 240
Db 181 CTGCGCTGCTGCCCTCTTTGGCGCGGGAAGCAGCACCAAGTTTTCACGGCCAAACGCTTTGGC 240
QY 241 ACTAGGGTCAGAAATGGCTTACAACAGTCCCTGATGTTGGCGCAATGGCCCTGAAATCAA 300
Db 241 ACTAGGGTCAGAAATGGCTTACAACAGTCCCTGATGTTGGCGCAATGGCCCTGAAATCAA 300
QY 301 GTACTACAGACTTTGTGATAAGCTTCAAGCTTCGGGCATCGTCTAGAAAACGGTGGCCAC 360
Db 301 GTACTACAGACTTTGTGATAAGCTTCAAGCTTCGGGCATCGTCTAGAAAACGGTGGCCAC 360
QY 361 AGCCGGGTTGTGAACCTCGGTGGCTTTCATGTCTCACTCTCCGATCCTCGTCTGCAAGGT 420
Db 361 AGCCGGGTTGTGAACCTCGGTGGCTTTCATGTCTCACTCTCCGATCCTCGTCTGCAAGGT 420
QY 421 GAGGACTCCAACAGCGGAAAATGCTGCTACTCAAGTTTCTTCTCTCTGCGGTGTGTT 480
Db 421 GAGGACTCCAACAGCGGAAAATGCTGCTACTCAAGTTTCTTCTCTCTGCGGTGTGTT 480
QY 481 GGGCATCTTTGGCTTCACTTCCCTTTCATCGGACTGCGGAGGAGCACAGGGCCAC 540
Db 481 GGGCATCTTTGGCTTCACTTCCCTTTCATCGGACTGCGGAGGAGCACAGGGCCAC 540
QY 541 AGCTTTCTTCTCTTTGGGATCCTCTTTTCCATCTCTCTCTCTGCTGCTCATGC 600
Db 541 AGCTTTCTTCTCTTTGGGATCCTCTTTTCCATCTCTCTCTCTGCTGCTCATGC 600
QY 601 TGTCACTTGACCAAGCTCGTCCGGGGAGGAAGCCCTTTTCTCTGTTGGTGTATTCGGG 660
Db 601 TGTCACTTGACCAAGCTCGTCCGGGGAGGAAGCCCTTTTCTCTGTTGGTGTATTCGGG 660
QY 661 TCTGGCCGTGGGCTTCAGCTAGTCCAGGATGTTATCGCTATTGAATATATTGCTCTGAC 720
Db 661 TCTGGCCGTGGGCTTCAGCTAGTCCAGGATGTTATCGCTATTGAATATATTGCTCTGAC 720
QY 721 CATGAATAGGACCAACGCTCAATGTCTTTTCTGAGCTTTTCGGCTCTCTCGTCAATGAAGA 780
Db 721 CATGAATAGGACCAACGCTCAATGTCTTTTCTGAGCTTTTCGGCTCTCTCGTCAATGAAGA 780
QY 781 CTTTGTCTCTCTGCTCACTAGCTCTCTCTTTGATGGGCTGACCTTCTCTATGCTCTC 840
Db 781 CTTTGTCTCTCTGCTCACTAGCTCTCTCTTTGATGGGCTGACCTTCTCTATGCTCTC 840
QY 841 CTTACCTTCTGTTGGTTCCTTCAACGGCTGGAAGACATGGGGCCCACTACCTCTAC 900
Db 841 CTTACCTTCTGTTGGTTCCTTCAACGGCTGGAAGACATGGGGCCCACTACCTCTAC 900
QY 901 GATGCTCTCTCCATTGCCATCTGGGTGGCTGGATCACCTGCTCATGCTTCTGACTT 960
Db 901 GATGCTCTCTCCATTGCCATCTGGGTGGCTGGATCACCTGCTCATGCTTCTGACTT 960
QY 961 TGACCGCAGGTGGGATGACCACTCTAGCTCCGGCTTGGCTGGCAATGGCTGGGTGTT 1020
Db 961 TGACCGCAGGTGGGATGACCACTCTAGCTCCGGCTTGGCTGGCAATGGCTGGGTGTT 1020
QY 1021 CTTGTTGGCTTATGTTAGTCCCGAGTTTGGCTGCTCACAAGCAACGAAACCCATGGA 1080
Db 1021 CTTGTTGGCTTATGTTAGTCCCGAGTTTGGCTGCTCACAAGCAACGAAACCCATGGA 1080
QY 1081 TTATCCTGTTGAGGATGCTTTCTGTAAACCTCAACTCGTGAAGAGAGCTATGCTGGA 1140
Db 1081 TTATCCTGTTGAGGATGCTTTCTGTAAACCTCAACTCGTGAAGAGAGCTATGCTGGA 1140
QY 1141 GAACAGAGCTACTCTCAAGAGAGAAATCACTCAAGGTTTTTGAAGAGACAGGGACAGCT 1200
Db 1141 GAACAGAGCTACTCTCAAGAGAGAAATCACTCAAGGTTTTTGAAGAGACAGGGACAGCT 1200

```



QY 1201 CTATGCCCCCTATTCCACACATTTTCAGCTGCAGAACCCAGCTCCCCCAAGGAATTTCTC 1260  
Db 1201 CTATGCCCCCTATTCCACACATTTTCAGCTGCAGAACCCAGCTCCCCCAAGGAATTTCTC 1260  
QY 1261 CATCCACAGGGCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAGAAAGAGGG 1320  
Db 1261 CATCCACAGGGCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAGAAAGAGGG 1320  
QY 1321 CAGCTAACTCTGTCTGTAAGAGTGGACAAATGACGCCGGCGGAGAGTCTAGCGGGAGC 1380  
Db 1321 CAGCTAACTCTGTCTGTAAGAGTGGACAAATGACGCCGGCGGAGAGTCTAGCGGGAGC 1380  
QY 1381 TCNAAAGGATGGGGGAAATCTTGAGTCTTCTGAGAAACTGTACAAGACACTACGGGA 1440  
Db 1381 TCNAAAGGATGGGGGAAATCTTGAGTCTTCTGAGAAACTGTACAAGACACTACGGGA 1440  
QY 1441 ACAGTTTGCTCCCTCCAGGCTCAACCAATTTCTTCATGCTGGGGCTGATGGGGCT 1500  
Db 1441 ACAGTTTGCTCCCTCCAGGCTCAACCAATTTCTTCATGCTGGGGCTGATGGGGCT 1500  
QY 1501 AGTAAGACTCCAGTTCTTAGAGCGCTGTAGTATTTTTTTTTTTTGTCTCATCTTTGG 1560  
Db 1501 AGTAAGACTCCAGTTCTTAGAGCGCTGTAGTATTTTTTTTTTTTGTCTCATCTTAG 1560  
QY 1561 ATACTTCTTTTAAGTGGGAGTCTCAGCAACTCAAGTTTAGACCTTACTCTTTTGT 1620  
Db 1561 ATACTTCTTTTAAGTGGGAGTCTCAGCAACTCAAGTTTAGACCTTACTCTTTTGT 1620  
QY 1621 GTTTTTTGAACAGGATCTTGCTCTGTACCCAGGCTTGAGTGCAGTGGCGCATCACAG 1680  
Db 1621 GTTTTTTGAACAGGATCTTGCTCTGTACCCAGGCTTGAGTGCAGTGGCGCATCACAG 1680  
QY 1681 CCAGTGCAGCTCGACCACTGTCTGAAGCAATCTCCATCTCCATCTCCCAAGTG 1740  
Db 1681 CCAGTGCAGCTCGACCACTGTCTGAAGCAATCTCCATCTCCATCTCCCAAGTG 1740  
QY 1741 CTGGGATGACAGCGTGAAGCAAGTCTCCAGCTAGGCCCTTAATCTTGTGTTATTTT 1800  
Db 1741 CTGGGATGACAGCGTGAAGCAAGTCTCCAGCTAGGCCCTTAATCTTGTGTTATTTT 1800  
QY 1801 CCATGACTTAAAGGCTTGGTCTCATGAGTCTCAGCTGGGTTCACACAGCTCTAGGGCCCTG 1860  
Db 1801 CCATGACTTAAAGGCTTGGTCTCATGAGTCTCAGCTGGGTTCACACAGCTCTAGGGCCCTG 1860  
QY 1861 CTCCTCTAATCTACAGTGGGTTTGTAGGCTCTGTGGCCAGAGCAGACCTGCATATCT 1920  
Db 1861 CTCCTCTAATCTACAGTGGGTTTGTAGGCTCTGTGGCCAGAGCAGACCTGCATATCT 1920  
QY 1921 GAGCAAAATAGCAAAAGCTCTCTCAGCCCACTGGCCTGAACTCTACACTGGAGCCAAAC 1980  
Db 1921 GAGCAAAATAGCAAAAGCTCTCTCAGCCCACTGGCCTGAACTCTACACTGGAGCCAAAC 1980  
QY 1981 TTGCTGGCACCCCGCTCCCAACCTTCTTGCTGGGTAGGAGGCTTAAAGATCACCC 2040  
Db 1981 TTGCTGGCACCCCGCTCCCAACCTTCTTGCTGGGTAGGAGGCTTAAAGATCACCC 2040  
QY 2041 TAAATTTACTCATCTCTAGTGTCTCCATCAATGGGCTCAGAGCTCCCGAGACCA 2100  
Db 2041 TAAATTTACTCATCTCTAGTGTCTCCATCAATGGGCTCAGAGCTCCCGAGACCA 2100  
QY 2101 ATTACAGGTACCCCTCTCTTCTGCACTGTCGCCAACTTGTCTCAATTCAGATC 2160  
Db 2101 ATTACAGGTACCCCTCTCTTCTGCACTGTCGCCAACTTGTCTCAATTCAGATC 2160  
QY 2161 TAAATCTCCCTTACTGCTGCAAGGAATCTTTTTCAGACCTCACTAGCAAAAGCCGGTTG 2220  
Db 2161 TAAATCTCCCTTACTGCTGCAAGGAATCTTTTTCAGACCTCACTAGCAAAAGCCGGTTG 2220  
QY 2221 CTCCTTGTGAGGAAATTTGTAGATCAATCTCACTTCAATTCCTGGGGCTGATATCTCT 2280  
Db 2221 CTCCTTGTGAGGAAATTTGTAGATCAATCTCACTTCAATTCCTGGGGCTGATATCTCT 2280  
QY 2281 CTCATCTTGCACCCCAACCTCTCTAAATAGATTTTACCGCATTTTACGGCTGCTTCTGTAA 2340

Db 2281 CTCATCTTGCACCCCAACCTCTGTAATAGATTTACCGCATTTTACGGCTGCTTCTGTAA 2340  
QY 2341 GTGGGCATGGTCTCTTAATGAGGAGTGTTCATTGTATATAAGTTATTACACCTGAGTAT 2400  
Db 2341 GTGGGCATGGTCTCTTAATGAGGAGTGTTCATTGTATATAAGTTATTACACCTGAGTAT 2400  
QY 2401 GCAATAAAGATGTGTGGCCACTCTTTTCATGTGTGGGAGCA 2443  
Db 2401 GCAATAAAGATGTGTGGCCACTCTTTTCATGTGTGGGAGCA 2443

RESULT 6  
AX877483  
LOCUS AX877483 2446 bp DNA linear PAT 17-DEC-2003  
DEFINITION Sequence 12388 from Patent EP1074617.  
ACCESSION AX877483  
VERSION AX877483.1 GI:40032219  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
MAMMALIA; EUTHERIA; EUARCHONTOLIRES; PRIMATES; CATARTHINI;  
HOMINIDAE; HOMO.

REFERENCE 1  
AUTHORS Oca,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,  
Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.  
TITLE Primers for synthesising full-length cDNA and their use  
JOURNAL Patent: EP 1074617-A 12388 07-FEB-2001;  
Research Association for Biotechnology (Jp)  
FEATURES  
Location/Qualifiers  
source  
1..2446  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
254..1327  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAE89746.1"  
/db\_xref="GI:40032220"

CDS  
1  
translation="NRRKMLPTQFLFLGLVGLIQLTFAFIIGLDSTGTPRFVFLFG  
MLTLPILVCKVQDSNRKMLPTQFLFLGLVGLIQLTFAFIIGLDSTGTPRFVFLFG  
ILFSCFCLLAHAVSLTKLVGRKPLSLVILGLAVGFLAVQDVIAI EIVILTMNR  
NVNFEISAPRNEVLTLLTYVLFLMALTFILMSFTFCGSGTGHKRGHAIYLTM  
LSIALIWMATLILMLPDEFDRWDDTLISALAAANGVFLAIYVSPFWLITKORND  
YVEBDAFCPKQLVKKSYGVENRAYSQBEITQGFETGDTLYAPYSTHFLQNPPOKE  
FSIPRAHAWPSYPKYDEVKKEGS"

ORIGIN  
Query Match 99.3%; Score 2439.8; DB 2; Length 2446;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 2441; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 ATAAACAGCATGAAGTCCGTGGAATAGGCGTGTCTCTCGACCCCTCCCC 60  
Db 1 ATAAACAGCATGAAGTCCGTGGAATAGGCGTGTCTCTCGACCCCTCCCC 60  
QY 61 TCCTTGTCCCTCTGCTCACCCCTCGCTCCCTCCCTCCGCGAGGCGCCCTTATA 120  
Db 61 TCCTTGTCCCTCTGCTCACCCCTCGCTCCCTCCCTCCGCGAGGCGCCCTTATA 120  
QY 121 ACNACTGCTCAGATGCGAGGCGGGATAGCTGTCCAAGGTCTCCCGCAGCACTGAGGAG 180  
Db 121 ACNACTGCTCAGATGCGAGGCGGGATAGCTGTCCAAGGTCTCCCGCAGCACTGAGGAG 180  
QY 181 CTCGCTGTGCTCCCTCTTTCGCGCGGGAAGACAGCACAAAGTTCCACGCCAAACGCCCTTGGC 240  
Db 181 CTCGCTGTGCTCCCTCTTTCGCGCGGGAAGACAGCACAAAGTTCCACGCCAAACGCCCTTGGC 240  
QY 241 ACTAGGGTCCAGATGGCTTACAACTGCTGATGTTGCGCAATGGCCCTGAAATCCAA 300  
Db 241 ACTAGGGTCCAGATGGCTTACAACTGCTGATGTTGCGCAATGGCCCTGAAATCCAA 300







```
Db 721 CATGAATAGGACCAACGTCATATGCTTTCTGAGCTTTCCGCTCCTCGTCGCAATGAAGA 780
Qy 781 CTTTGTCTCTGCTGCTACCTACGTCCTCTCTCTTCTGATCGCGCTGACCTTTCTCTCATGTCTCTC 840
Db 781 CTTTGTCTCTGCTGCTACCTACGTCCTCTCTCTTCTGATGGCGTGACCTTTCTCTCATGTCTCTC 840
Qy 841 CTTTCACTCTCTGTTGTTCTTTTCTGCGGCTGGAAGAGACATGGGGCCCACTACCTCAC 900
Db 841 CTTTCACTCTCTGTTGTTCTTTTCTGCGGCTGGAAGAGACATGGGGCCCACTACCTCAC 900
Qy 901 GATGCTCTCTCCATTTGCGCATCTGGGTGGCTGTGATCACCTGTCTCATGTCTCTGACTT 960
Db 901 GATGCTCTCTCTCCATTTGCGCATCTGGGTGGCTGTGATCACCTGTCTCATGTCTCTGACTT 960
Qy 961 TGACCGGAGGTGGGATGACACATCTCTCAGCTCGCGCTTGGCTGCGCAATGCGTGGGTGTT 1020
Db 961 TGACCGGAGGTGGGATGACACATCTCTCAGCTCGCGCTTGGCTGCGCAATGCGTGGGTGTT 1020
Qy 1021 CTTGTTGGCTTATGTTAGTCCCGAGTTTGGCTGCTCACAAGCAACGAAACCCCATGGA 1080
Db 1021 CTTGTTGGCTTATGTTAGTCCCGAGTTTGGCTGCTCACAAGCAACGAAACCCCATGGA 1080
Qy 1081 TTATCTCTGTTGAGATGCTTTCTGTAACCTCAACTCGTGAAGAGACTATGTTGGA 1140
Db 1081 TTATCTCTGTTGAGATGCTTTCTGTAACCTCAACTCGTGAAGAGACTATGTTGTTGA 1140
Qy 1141 GAACAGAGCCTACTCTCAAGAGGAATCACTCAAGGTTTGAAGAGACAGGGACAGCT 1200
Db 1141 GAACAGAGCCTACTCTCAAGAGGAATCACTCAAGGTTTGAAGAGACAGGGACAGCT 1200
Qy 1201 CTATGCCCTTATTTCCACACATTTTTCAGCTGCGAAGCCAGCTCCCCAAAAGGAATTTCTC 1260
Db 1201 CTATGCCCTTATTTCCACACATTTTTCAGCTGCGAAGCCAGCTCCCCAAAAGGAATTTCTC 1260
Qy 1261 CATCCACGGGCGCCACGCTGGCGGAGCCCTTACAAAGACTATGAAGTAAAGAAAGGG 1320
Db 1261 CATCCACGGGCGCCACGCTGGCGGAGCCCTTACAAAGACTATGAAGTAAAGAAAGGG 1320
Qy 1321 CAGCTAACTCTGTCTGAAGAGTGGGACAAATGACGCGGCGGCGAGATCTAGCGGGAGC 1380
Db 1321 CAGCTAACTCTGTCTGAAGAGTGGGACAAATGACGCGGCGGCGAGATCTAGCGGGAGC 1380
Qy 1381 TCAAGGAGTGTGGCGGAATTTTGTAGTCTTCTGAGAAAACCTGTACAAGACACTACGGGA 1440
Db 1381 TCAAGGAGTGTGGCGGAATTTTGTAGTCTTCTGAGAAAACCTGTACAAGACACTACGGGA 1440
Qy 1441 ACAGTTTGCTCTCCCTCCAGCTCAACCAATTTCTCCATGCTGGGGCTGATGGGCT 1500
Db 1441 ACAGTTTGCTCTCCCTCCAGCTCAACCAATTTCTCCATGCTGGGGCTGATGGGCT 1500
Qy 1501 AGTAAGACTCCAGTTCTTAGAGCGCTGTAGTATTTTTTTTTTTTTTTTGTCTCATCCTTGG 1560
Db 1501 AGTAAGACTCCAGTTCTTAGAGCGCTGTAGTATTTTTTTTTTTTTTTTGTCTCATCCTTGG 1560
Qy 1561 ATACTTTCTTTTAAAGTGGAGTCTCAGGCAACTCAAGTTTACAGCCCTTACTCTTTTGT 1620
Db 1561 ATACTTTCTTTTAAAGTGGAGTCTCAGGCAACTCAAGTTTACAGCCCTTACTCTTTTGT 1620
Qy 1621 GTTTTTGAAACAGGATCTTTGCTCTGTCAACCGAGCTTGTAGTGCAGTGGGATCACAG 1680
Db 1621 GTTTTTGAAACAGGATCTTTGCTCTGTCAACCGAGCTTGTAGTGCAGTGGGATCACAG 1680
Qy 1681 CCAGTGCAGCTCGACCTGTGCTCAGCAATCTCCCATCTCCATCTCCCATCTCCCAAGTG 1740
Db 1681 CCAGTGCAGCTCGACCTGTGCTCAGCAATCTCCCATCTCCATCTCCCATCTCCCAAGTG 1740
Qy 1741 CTGGATGACAGGCGTGAGCCACAGCTCCAGCTAGGCGCTTAATCTTCTGCTGTTATTT 1800
Db 1741 CTGGATGACAGGCGTGAGCCACAGCTCCAGCTAGGCGCTTAATCTTCTGCTGTTATTT 1800
Qy 1801 CCATGACTAAAGGTCTGGTCTATCTGAGCTCAGCTGGCTCACACAGCTCTAGGGCGCTG 1860
Db 1801 CCATGACTAAAGGTCTGGTCTATCTGAGCTCAGCTGGCTCACACAGCTCTAGGGCGCTG 1860
```

```
Qy 1861 CTCTCTTAACCTCACAGTGGGTTTTTGTGAGGCTCTGTGGCCGAGAGACAGCTGCATATCT 1920
Db 1861 CTCTCTTAACCTCACAGTGGGTTTTTGTGAGGCTCTGTGGCCGAGAGACAGCTGCATATCT 1920
Qy 1921 GAGCAAAAATAGCAAAAGCCCTCTCTCAGCCCACTGGCCTGAAATCTACACTGGAAGCCAAC 1980
Db 1921 GAGCAAAAATAGCAAAAGCCCTCTCTCAGCCCACTGGCCTGAAATCTACACTGGAAGCCAAC 1980
Qy 1981 TTGCTGGCACCCTGGCTCCCAACCCCTTCTTGTGCTGGTAGGAGAGGCTAAAGATCACCC 2040
Db 1981 TTGCTGGCACCCTGGCTCCCAACCCCTTCTTGTGCTGGTAGGAGAGGCTAAAGATCACCC 2040
Qy 2041 TAAATTTACTCATCTCTCTAGTGTGCTCACAATTTGGGCTCAGCAGCTCCCCAGCACCA 2100
Db 2041 TAAATTTACTCATCTCTCTAGTGTGCTCACAATTTGGGCTCAGCAGCTCCCCAGCACCA 2100
Qy 2101 ATTACAGGTCACCCCTCTCTTCTGACATGTCCCCAACTTCTGTCAATTCGAGATC 2160
Db 2101 ATTACAGGTCACCCCTCTCTTCTGACATGTCCCCAACTTCTGTCAATTCGAGATC 2160
Qy 2161 TAATCTCCCTTACGCTCTGCGAGGAATTTCTTTTTCAGACCTCAGTACGACAGCCGGTTG 2220
Db 2161 TAATCTCCCTTACGCTCTGCGAGGAATTTCTTTTTCAGACCTCAGTACGACAGCCGGTTG 2220
Qy 2221 CTCCTTGTGAGGAGAAATTTGTAGATCATTTCTCACTTCAAAATTCCTGGGGCTGATCTTCT 2280
Db 2221 CTCCTTGTGAGGAGAAATTTGTAGATCATTTCTCACTTCAAAATTCCTGGGGCTGATCTTCT 2280
Qy 2281 CTCATCTTGCACCCCAACCTCTGTAATAGATTTACCGCATTTACGGCTCATTTCTGTAA 2340
Db 2281 CTCATCTTGCACCCCAACCTCTGTAATAGATTTACCGCATTTACGGCTCATTTCTGTAA 2340
Qy 2341 GTGGGATGCTCTCTTAATGGAGAGTGTTCATTTGATATTAATAGTTATTACCTGAGTAT 2400
Db 2341 GTGGGATGCTCTCTTAATGGAGAGTGTTCATTTGATATTAATAGTTATTACCTGAGTAT 2400
Qy 2401 GCAATAAAGATGTGGTGGCCACTCTTTTCATGTTGGTGGCAGCA 2443
Db 2401 GCAATAAAGATGTGGTGGCCACTCTTTTCATGTTGGTGGCAGCA 2443
```

RESULT 8  
AK122672

LOCUS AK122672 3057 bp mRNA linear PRI 20-JAN-2006  
DEFINITION Homo sapiens cDNA FLJ16117 fis, clone ASTR02003632, highly similar to Homo sapiens putative G protein-coupled receptor (RAIG1) mRNA.

ACCESSION AK122672

VERSION AK122672.1 GI:34527861

KEYWORDS oligo capping; fis (full insert sequence).

SOURCE Homo sapiens

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

## AUTHORS

Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R., Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H., Sekine, M., Oobayashi, M., Nishi, T., Shibahara, T., Tanaka, T., Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y., Nagahara, K., Murakami, K., Yasuda, T., Iwayanagi, T., Wagatsuma, M., Shiratori, A., Sudo, H., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H., Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T., Kikkawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K., Tanikawa, M., Yamazaki, M., Ninomiya, K., Ishibashi, T., Yamashita, H., Murakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M., Hiraoka, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S., Yosida, M., Horiuta, T., Kusano, J., Kanehori, K., Takahashi-Fujii, A., Hara, H., Tanase, T., O., Nomura, Y., Togiya, S., Komai, F., Hara, R., Takeuchi, K., Arita, M., Imose, N., Musashino, K., Yuuki, H., Ohnita, A., Sasaki, N., Aotaka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T., Shiohata, N., Sano, S., Moriya, S., Momiyama, H., Satoh, N., Takami, S., Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H.,







```
/db xref="taxon:9606"
/chrmosome="12"
/map="12p12.3-p13"
1. .2302
/genes="RAIG1"
/notes="induced by retinoic acid"
100. .1173
/genes="RAIG1"
/codon_start=1
/product="putative G protein-coupled receptor"
/protein_id="AAC98506.1"
/db xref="GI:4063890"
/translation="MATTVPDGRNGLKSYRLCDKABAWGIVLETVATGAVTVSA
FMILTPLVCKQDSNRKMLPTQFLFLGLVIGLTFAPFIIGLDSGTPFPFLFG
ILFSICFSLLAHVSLLKLVGRKPLSLVILGLAVGSLVQDVIAISYIVLTMRT
NVNFSBLSAPRNEFDVLLTYVLFMALTYFLMSSFTFCGSGFTGWRKHGAIYLLML
LSIAIWMATVILMLPDRDDTILSSALAAWGVFLAYVSPFLLTTRKORPMD
YVEDAFCKPLVKSYGVENRAYSQEIEITQGEETGDTLYAPYSTHFLQNPQPOKE
FSIPRAHAWSPSPYKDYEVKKEGS"
ORIGIN
Query Match 93.7%; Score 2302; DB 5; Length 2302;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2302; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 155 CCAAGTCTCCCGACGACTGAGGAGCTCGCCTGCTGCCCTCTTGGCGGGGGAAGCAGC 214
DB 1 CCAAGTCTCCCGACGACTGAGGAGCTCGCCTGCTGCCCTCTTGGCGGGGGAAGCAGC 60
QY 215 ACCAAGTCTCAAGGCCCAACGCTTGGCACTAGGCTCCAGAAATGCTGCTCAACAGCTCCCTGAT 274
DB 61 ACCAAGTCTCAAGGCCCAACGCTTGGCACTAGGCTCCAGAAATGCTGCTCAACAGCTCCCTGAT 120
QY 275 GGTGGCGCAATGGCTGAAATCCAAAGTACTACAGACTTTGTGATGAAGCTGAAGCTTG 334
DB 121 GGTGGCGCAATGGCTGAAATCCAAAGTACTACAGACTTTGTGATGAAGCTGAAGCTTG 180
QY 335 GGCATCGTCTAGAAAGCTGGCCACAGCGGGGTTGTGACCTCGGTGGCTTCATGCTC 394
DB 181 GGCATCGTCTAGAAAGCTGGCCACAGCGGGGTTGTGACCTCGGTGGCTTCATGCTC 240
QY 395 ACTCTCCGATCTCGTCTGCAAGTGCAGGACTCCAAACAGGCGGAAATGCTGCCTACT 454
DB 241 ACTCTCCGATCTCGTCTGCAAGTGCAGGACTCCAAACAGGCGGAAATGCTGCCTACT 300
QY 455 CAGTTTCTCTCTCGGTGTTGGGCACTTTGGGCTACCTTGGCTTCGCTTCATCATC 514
DB 301 CAGTTTCTCTCTCGGTGTTGGGCACTTTGGGCTACCTTGGGCTTCGCTTCATCATC 360
QY 515 GGAATGGAGGAGCAGCGGCCACACGCTTCTCTTTGGGATCCTCTTTTCCATC 574
DB 361 GGAATGGAGGAGCAGCGGCCACACGCTTCTCTTTGGGATCCTCTTTTCCATC 420
QY 575 TGCTTCTCTCTCGTCTGCTCATGTGTGCTGACCAAGCTCGTCCGGGGGAGGAG 634
DB 421 TGCTTCTCTCTCGTCTGCTCATGTGTGCTGACCAAGCTCGTCCGGGGGAGGAG 480
QY 635 CCCCTTCTCTGTTGTGATTTGGGCTTGGCGGCTTACGCTTACGCTTACAGGATGT 694
DB 481 CCCCTTCTCTGTTGTGATTTGGGCTTGGCGGCTTACGCTTACGCTTACAGGATGT 540
QY 695 ATCGCTATTGAATATATGCTGACCATGATAGGACCAAGCTCAATGCTTTCTGAG 754
DB 541 ATCGCTATTGAATATATGCTGACCATGATAGGACCAAGCTCAATGCTTTCTGAG 600
QY 755 CTTTCCGCTCTCGTCCGAATGAAGACTTTGCTCTGCTCACTACGCTCTCTTTCTTG 814
DB 601 CTTTCCGCTCTCGTCCGAATGAAGACTTTGCTCTGCTCACTACGCTCTCTTTCTTG 660
QY 815 ATGGCGCTGACCTTCTCATGCTCTCTTCACTTCTGTGGTTCCTTACGGGCTGGAAG 874
DB 661 ATGGCGCTGACCTTCTCATGCTCTCTTCACTTCTGTGGTTCCTTCACTGCTGGAAG 720
```

```
QY 875 AGACATGGGGCCACATCTACCTCAGATGCTCTCTCATTTGCCATCTGGGTGGCTGG 934
DB 721 AGACATGGGGCCACATCTACCTCAGATGCTCTCTCATTTGCCATCTGGGTGGCTGG 780
QY 935 ATCACCTCTGCTCATGCTTCTGACTTTGACCGCAGGTGGGATGACACCATCTCAGCTCC 994
DB 781 ATCACCTCTGCTCATGCTTCTGACTTTGACCGCAGGTGGGATGACACCATCTCAGCTCC 840
QY 995 GCCTTGGCTGCCAATGGCTGGGTGTTCTGTTGGCTTATGTTAGTCCGAGATTTGGCTG 1054
DB 841 GCCTTGGCTGCCAATGGCTGGGTGTTCTGTTGGCTTATGTTAGTCCGAGATTTGGCTG 900
QY 1055 CTCACAAAAGCAACGAAACCCATGGAATATCTGTTGGATGCTTTCTGTAACCTCAA 1114
DB 901 CTCACAAAAGCAACGAAACCCATGGAATATCTGTTGAGGATGCTTTCTGTTAAACCTCAA 960
QY 1115 CTCGTGAAGAAGAGCTATGTTGGGAGAAACAGAGCCTACTCTCAAGAGGAAATCACTCAA 1174
DB 961 CTCGTGAAGAAGAGCTATGTTGGGAGAAACAGAGCCTACTCTCAAGAGGAAATCACTCAA 1020
QY 1175 GGTTTTGAAGAGACAGGGGACAGCTCTATGCCCTTATTCACACATTTTCAGCTGCAG 1234
DB 1021 GGTTTTGAAGAGACAGGGGACAGCTCTATGCCCTTATTCACACATTTTCAGCTGCAG 1080
QY 1235 AACCAGCTCCCCCAAAAGGAATTTCTCCATCCACGGGCCACGCTTGGCCGAGCCCTTAC 1294
DB 1081 AACCAGCTCCCCCAAAAGGAATTTCTCCATCCACGGGCCACGCTTGGCCGAGCCCTTAC 1140
QY 1295 AAAGACTATGAAGTAAAGAAAGAGGAGGAGCTAACTGTCTCTGAAGAGTGGGCAAAATGC 1354
DB 1141 AAAGACTATGAAGTAAAGAAAGAGGAGGAGCTAACTGTCTCTGAAGAGTGGGCAAAATGC 1200
QY 1355 AGCGGGCGGAGAGTCTAGCGGAGCTCAAGGGATGTGGCGGAAATCTTGTAGTCTTCTG 1414
DB 1201 AGCGGGCGGAGAGTCTAGCGGAGCTCAAGGGATGTGGCGGAAATCTTGTAGTCTTCTG 1260
QY 1415 AGAAACTCTACAAGACACTACGGGAAACAGTTTGGCTCCTCCCGACCTCAACACAAAT 1474
DB 1261 AGAAACTCTACAAGACACTACGGGAAACAGTTTGGCTCCTCCCGACCTCAACACAAAT 1320
QY 1475 CTTCCATGTGGGGTGTATGTGGGTAGTAAGACTCCAGTTCTTAGAGCGCTGTAGTAT 1534
DB 1321 CTTCCATGTGGGGTGTATGTGGGTAGTAAGACTCCAGTTCTTAGAGCGCTGTAGTAT 1380
QY 1535 TTTTCTTTTCTCTCATCTTGGATCTTTTAAAGAGGATCTTGTAGTGGGAGTCTCAGGCAACTCA 1594
DB 1381 TTTTCTTTTCTCTCATCTTGGATCTTTTAAAGAGGATCTTGTAGTGGGAGTCTCAGGCAACTCA 1440
QY 1595 AGTTTAGACCTTACTCTTTTGTGTTTGTGAAACAGGATCTTGTCTGTCAACCAG 1654
DB 1441 AGTTTAGACCTTACTCTTTTGTGTTTGTGAAACAGGATCTTGTCTGTCAACCAG 1500
QY 1655 GCTTGAGTGCAGTGTGGATCACAGCCAGTGCAGCTCGACCTCGACCTCTGTCTCAAGCAA 1714
DB 1501 GCTTGAGTGCAGTGTGGATCACAGCCAGTGCAGCTCGACCTCGACCTCTGTCTCAAGCAA 1560
QY 1715 TCCTCCATCTCATCTCCAAAGTCTGGGATGACAGGGGTGAGCCAGCTCCAGCC 1774
DB 1561 TCCTCCATCTCATCTCCAAAGTCTGGGATGACAGGGGTGAGCCAGCTCCAGCC 1620
QY 1775 TAGGCCCTTAATCTTGTCTGTTATTTTCCATGAAGTCTGCTCATCTGAGCTCACG 1834
DB 1621 TAGGCCCTTAATCTTGTCTGTTATTTTCCATGAAGTCTGCTCATCTGAGCTCACG 1680
QY 1835 CTGGCTCACACAGCTCTTAGGGCTGCTCTCTTAATCTCAGTGGGTTTTGTGAGGCTCT 1894
DB 1681 CTGGCTCACACAGCTCTTAGGGCTGCTCTCTTAATCTCAGTGGGTTTTGTGAGGCTCT 1740
QY 1895 GTGGCCGAGAGAGCTCTGCATATCTGAGCAAAAATAGCAAAAGCTCTCTCAGCCACT 1954
DB 1741 GTGGCCGAGAGAGCTCTGCATATCTGAGCAAAAATAGCAAAAGCTCTCTCAGCCACT 1800
QY 1955 GGCCTGAATCTACACTGGAAGCAACTTGTCTGCGACCCCGCTCCCAACCCCTTCTTGCC 2014
```







1257	Db		GGGCAAAATG	CAGCGGG	CGCAGATCTAG	CGGAGCTCAA	AGGGAATG	TGGGGAAATCT	1316
1404	Qy		TGAGCTTCT	GAGAAAACTGT	ACAAGACACTAC	GGGAACAGTTT	TGGCTCCT	CCCCAGCCT	1463
1317	Db		TGAGCTTCT	GAGAAAACTGT	ACAAGACACTA	CGGGAACAGTTT	TGGCTCCT	CCCCAGCCT	1376
1464	Qy		CAACCAAAAT	CTTCCATGCT	TGGGGCTGAT	TGGGCTAGTAA	AGACTCCAGT	CTTTAGAGG	1523
1377	Db		CAACCAAAAT	CTTCCATGCT	TGGGGCTGAT	TGGGCTAGTAA	AGACTCCAGT	CTTTAGAGG	1436
1524	Qy		CGCTGTAGT	ATTTTTTTTT	TTTGTCTCAT	CTCTTTTGGAT	ACTCTTTTAA	GTGGGAGTCT	1583
1437	Db		CGCTGTAGT	ATTTTTTTTT	TTTGTCTCAT	CTCTTTTGGAT	ACTCTTTTAA	GTGGGAGTCT	1496
1584	Qy		CAGGCAACT	CAAGTTT	TAGACCTT	PACTCTTTT	TGTTTGTAA	AAACAGGATCTTGCT	1643
1497	Db		CAGGCAACT	CAAGTTT	TAGACCTT	PACTCTTTT	TGTTTGTAA	AAACAGGATCTTGCT	1556
1644	Qy		CTGTCAACCC	AGGCTTGAGT	GCAGTGTG	CGATCACAG	CCCGAGTGCAG	CCCTCGACCACTG	1703
1557	Db		CTGTCAACCC	AGGCTTGAGT	GCAGTGTG	CGATCACAG	CCCGAGTGCAG	CCCTCGACCACTG	1616
1704	Qy		TGCTCAAGCA	ATCCTCCCAT	CTCCATCTCCAT	CTCCAAAGT	CTGGATGACAG	CGGTGAGCCAC	1763
1617	Db		TGCTCAAGCA	ATCCTCCCAT	CTCCATCTCCAT	CTCCAAAGT	CTGGATGACAG	CGGTGAGCCAC	1676
1764	Qy		AGCTCCAG	CCTAGGCCCTTAA	CTTGCTGTATTTT	TCATGGACTAA	AGGCTCTGGT	TCAT	1823
1677	Db		AGCTCCAG	CCTAGGCCCTTAA	CTTGCTGTATTTT	TCATGGACTAA	AGGCTCTGGT	TCAT	1736
1824	Qy		CTGAGCTC	ACGCTGGCTC	CACACAGTCT	TAGGGGCTGCT	CTCTCTAA	CTCACAGTGGGTTT	1883
1737	Db		CTGAGCTC	ACGCTGGCTC	CACACAGTCT	TAGGGGCTGCT	CTCTCTAA	CTCACAGTGGGTTT	1796
1884	Qy		TGTGAGGCT	CTGTGGCC	CAGACAGACCT	TGCATATCTG	AGCAAAAT	TAGCAAAAGCCTCT	1943
1797	Db		TGTGAGGCT	CTGTGGCC	CAGACAGACCT	TGCATATCTG	AGCAAAAT	TAGCAAAAGCCTCT	1856
1944	Qy		CTCAGCCACT	GGCTCGAATCT	ACACTGGA	AGCCACTGCT	GGCACCCCGCT	CCCCAA	2003
1857	Db		CTCAGCCACT	GGCTCGAATCT	ACACTGGA	AGCCACTGCT	GGCACCCCGCT	CCCCAA	1916
2004	Qy		CCCTTCTTG	CTGGT	TAGGAGGCTAA	AGATCACCCCTAA	ATTTACTCAT	CTCTCTAGTG	2063
1917	Db		CCCTTCTTG	CTGGT	TAGGAGGCTAA	AGATCACCCCTAA	ATTTACTCAT	CTCTCTAGTG	1976
2064	Qy		CTGCTCACA	ATTGGGCT	CTCAGCAGTCCC	CAGCAACAA	ATTACAGG	TCACCCCTCTCTTC	2123
1977	Db		CTGCTCACA	ATTGGGCT	CTCAGCAGTCCC	CAGCAACAA	ATTACAGG	TCACCCCTCTCTTC	2036
2124	Qy		TTGCACTG	CTCCCAACT	CTGCTCA	ATTCCGAGATCTA	ATCTCCCCCT	TAGCTCTGCCA	2183
2037	Db		TTGCACTG	CTCCCAACT	CTGCTCA	ATTCCGAGATCTA	ATCTCCCCCT	TAGCTCTGCCA	2096
2184	Qy		GGAAATCT	TTTTCAGAC	CTCATTAG	CACAAGCCCGG	TGCTCTCTT	TGTACAGGAAATTTGTAG	2243
2097	Db		GGAAATCT	TTTTCAGAC	CTCATTAG	CACAAGCCCGG	TGCTCTCTT	TGTACAGGAAATTTGTAG	2156
2244	Qy		ATCATTTCT	CAATTCAAA	ATTCTGGG	GTGATATCTTCT	CTCATCTTG	CAACCCCACTCTTG	2303
2157	Db		ATCATTTCT	CAATTCAAA	ATTCTGGG	GTGATATCTTCT	CTCATCTTG	CAACCCCACTCTTG	2216
2304	Qy		TAAATAGAT	TTACCGCAT	TTTACGGCTGC	ATTCTGTAG	TGGGCATGGT	CTCTTAATGGAG	2363
2217	Db		TAAATAGAT	TTACCGCAT	TTTACGGCTGC	ATTCTGTAG	TGGGCATGGT	CTCTTAATGGAG	2276
2364	Qy		GAGTGTTC	ATTGATATA	AGTTTATT	CACCTGAGTAT	GCAATAA	AGATGTGGTGGCACT	2423
2277	Db		GAGTGTTC	ATTGATATA	AGTTTATT	CACCTGAGTAT	GCAATAA	AGATGTGGTGGCACT	2336
2424	Qy		CTTTTCAT	GGTGGG	CAGCA	2443			

Db	2337 CTTTCATGGTGGCAGCA 2356									
RESULT 12										
CQ723177										
LOCUS	CQ723177	2290 bp	DNA	linear	PAT 03-FEB-2004					
DEFINITION	Sequence 9111 from Patent WO02068579.									
ACCESSION	CQ723177									
VERSION	CQ723177.1	GI:42284034								
KEYWORDS										
SOURCE	Homo sapiens (human)									
ORGANISM	Homo sapiens									
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.									
REFERENCE	1 Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.									
AUTHORS	Kits, such as nucleic acid arrays, comprising a majority of									
TITLE	humanexons or transcripts, for detecting expression and other uses thereof									
JOURNAL	Patent: WO 02068579-A 9111 06-SEP-2002;									
FEATURES	PE Corporation (NY) (US)									
source	Location/Qualifiers									
	1..2290									
	/organism="Homo sapiens"									
	/mol_type="unassigned DNA"									
	/db_xref="taxon:9606"									
ORIGIN										
Query Match	92.6%; Score 2274, 8; DB 2; Length 2290;									
Best Local Similarity	99.9%; Pred. No. 0;									
Matches 2287;	Conservative 0; Mismatches 2; Indels 1; Gaps 1;									
QY	155	CCAAGGTC	CCCCAGAC	TGAGAG	CTGCCCTG	CTGCCCCCTTTG	CGCGGGAAG	CAGC	214	
Db	1	CCAAGGTC	CCCCAGAC	TGAGAG	CTGCCCTG	CTGCCCCCTTTG	CGCGGGAAG	CAGC	60	
QY	215	ACCAAGTTC	ACGGCCAA	CGCCTT	GGCACTAG	GGTCCAGAA	TGGCTACAA	CAGTCCCTGAT	274	
Db	61	ACCAAGTTC	ACGGCCAA	CGCCTT	GGCACTAG	GGTCCAGAA	TGGCTACAA	CAGTCCCTGAT	120	
QY	275	GTTTGC	CGCAATGG	CCCTGAAAT	CAAGTACT	ACAGACTT	TGTGATAAG	GCTGAAGCTTGG	334	
Db	121	GTTTGC	CGCAATGG	CCCTGAAAT	CAAGTACT	ACAGACTT	TGTGATAAG	GCTGAAGCTTGG	180	
QY	335	GGCATG	CTCTAG	AAACGG	TGGCCAC	AGCGGGT	TGTGAC	CTCGTGGCCCTCATGCTC	394	
Db	181	GGCATG	CTCTAG	AAACGG	TGGCCAC	AGCGGGT	TGTGAC	CTCGTGGCCCTCATGCTC	240	
QY	395	ACTCTCC	CGATCCT	CGTCTC	GAAGTGC	AGGATCC	CAACAGG	CGGAAATATGCTGCCTACT	454	
Db	241	ACTCTCC	CGATCCT	CGTCTC	GAAGTGC	AGGATCC	CAACAGG	CGGAAATATGCTGCCTACT	300	
QY	455	CAGTTTCT	CTTCTCT	CGGTGTG	TGGG	GATCTTT	TGGCCTCAC	CTTTCGCTTTCATCATC	514	
Db	301	CAGTTTCT	CTTCTCT	CGGTGTG	TGGG	GATCTTT	TGGCCTCAC	CTTTCGCTTTCATCATC	360	
QY	515	GGACTG	ACGGGAG	CACAGG	GGCCAC	ACGCTTCT	TCTCTTTGGGAT	CTCTTTTCCATC	574	
Db	361	GGACTG	ACGGGAG	CACAGG	GGCCAC	ACGCTTCT	TCTCTTTGGGAT	CTCTTTTCCATC	420	
QY	575	TGCTTTC	CTGCTG	CTGGTCT	CACTGTG	TCAGTCT	GCACCAAG	CTCGTCCGGGGAGGAAG	634	
Db	421	TGCTTTC	CTGCTG	CTGGTCT	CACTGTG	TCAGTCT	GCACCAAG	CTCGTCCGGGGAGGAAG	480	
QY	635	CCCCTTT	CCCTGT	TGGTGA	TCTTGG	GTCTGG	CGTGGGCTTC	AGCCTAGTCCAGGATGTT	694	
Db	481	CCCCTTT	CCCTGT	TGGTGA	TCTTGG	GTCTGG	CGTGGGCTTC	AGCCTAGTCCAGGATGTT	540	
QY	695	ATCGCTAT	TGAAATAT	TGTTCCT	GACCAT	GAATAG	GCACCAAC	GTCAATGCTCTTTCTGAG	754	
Db	541	ATCGCTAT	TGAAATAT	TGTTCCT	GACCAT	GAATAG	GCACCAAC	GTCAATGCTCTTTCTGAG	600	

QY 755 CTTTCGGCTCCTCGTGCATGAAGACTTTGTGCTCCTCGCTCACCTACGTCTCTCTTTG 814  
DB 601 CTTTCGGCTCCTCGTGCATGAAGACTTTGTGCTCCTCGCTCACCTACGTCTCTCTTTG 660  
QY 815 ATGGGGCTGACCTTCTCATGTCCTCTCTCATCTTCTGTGGTTCCTTACAGGGCTGGAG 874  
DB 661 ATGGGGCTGACCTTCTCATGTCCTCTCTCATCTTCTGTGGTTCCTTACAGGGCTGGAG 720  
QY 875 AGACATGGGGCCACATCTACCTACGATGCTCTCTCCATTGCCATCTGGGTGGCCTGG 934  
DB 721 AGACATGGGGCCACATCTACCTACGATGCTCTCTCCATTGCCATCTGGGTGGCCTGG 780  
QY 935 ATCACCTGCTCATGCTTCTGACTTTGACCCAGGTGGGATGACACCATCTCAGCTCC 994  
DB 781 ATCACCTGCTCATGCTTCTGACTTTGACCCAGGTGGGATGACACCATCTCAGCTCC 840  
QY 995 GCCTTGGCTGCCAATGGCTGGGTGTTCTGTGGCTTATGTTAGTCCCGAGTTTGGCTG 1054  
DB 841 GCCTTGGCTGCCAATGGCTGGGTGTTCTGTGGCTTATGTTAGTCCCGAGTTTGGCTG 900  
QY 1055 CTCACAAAGCAACGAAACCCCATGATTAATCCTGTTGAGGATGCTTTCTGTAACCTCAA 1114  
DB 901 CTCACAAAGCAACGAAACCCCATGATTAATCCTGTTGAGGATGCTTTCTGTAACCTCAA 960  
QY 1115 CTCGTGAAGAAGACTATGGTGTGAGAACAGAGCTACTCTCAAGAGGAATCACTCA 1174  
DB 961 CTCGTGAAGAAGACTATGGTGTGAGAACAGAGCTACTCTCAAGAGGAATCACTCA 1020  
QY 1175 GGTGTTTGAAGACAGGGGACACGCTCTATGCCCTATTTCCACACATTTTCAGCTGCAG 1234  
DB 1021 GGTGTTTGAAGACAGGGGACACGCTCTATGCCCTATTTCCACACATTTTCAGCTGCAG 1080  
QY 1235 AACCGCCTCCCAAAAGGAATTTCCATCCACGCGGCCACGCTTGGCCGAGCCCTTAC 1294  
DB 1081 AACCGCCTCCCAAAAGGAATTTCCATCCACGCGGCCACGCTTGGCCGAGCCCTTAC 1140  
QY 1295 AAAGACTATGAAGTAAGAAAGGGCAGCTAACTGTCTGCTGAAGAGTGGGACAAATGC 1354  
DB 1141 AAAGACTATGAAGTAAGAAAGGGCAGCTAACTGTCTGCTGAAGAGTGGGACAAATGC 1200  
QY 1355 AGCCGGCGGCGAGATCTAGCGGGAGCTCAAGGGATGTGGCGAAATCTTGAGTCTTCG 1414  
DB 1201 AGCCGGCGGCGAGATCTAGCGGGAGCTCAAGGGATGTGGCGAAATCTTGAGTCTTCG 1260  
QY 1415 AGAAACTGTACAAGACACTACGGGAAGAGTTTGCTCCCTCCAGCCTCAACCAAAAT 1474  
DB 1261 AGAAACTGTACAAGACACTACGGGAAGAGTTTGCTCCCTCCAGCCTCAACCAAAAT 1320  
QY 1475 CTTCCATGCTGGGCTGATGTGGCTAGTAAGACTCCAGTCTTAGAGCGCTGTAGTAT 1534  
DB 1321 CTTCCATGCTGGGCTGATGTGGCTAGTAAGACTCCAGTCTTAGAGCGCTGTAGTAT 1380  
QY 1535 TTTTCTTTTGTCTCATCTCTTGATACTTCTTTTAAGTGGGAGTCTCAGGCAACTCA 1594  
DB 1381 TTTTCTTTTGTCTCATCTCTTGATACTTCTTTTAAGTGGGAGTCTCAGGCAACTCA 1440  
QY 1595 AGTTAGACCCCTACTCTTTTGTGTTTTTGAACAGGATCTTGCTGTACCCAG 1654  
DB 1441 AGTTAGACCCCTACTCTTTTGTGTTTTTGAACAGGATCTTGCTGTACCCAG 1500  
QY 1655 GCTTGAGTGCACTGGTGGCATCACAGCCAGTGCAGCTCGACCACTGCTGTCTCAGCAA 1714  
DB 1501 GCTTGAGTGCACTGGTGGCATCACAGCCAGTGCAGCTCGACCACTGCTGTCTCAGCAA 1560  
QY 1715 TCCTCCCATCTCCATCTCCCAAGTGTCTGGGATGACAGCGCTGAGCCACAGCTCCAGCC 1774  
DB 1561 TCCTCCCATCTCCATCTCCCAAGTGTCTGGGATGACAGCGCTGAGCCACAGCTCCAGCC 1620  
QY 1775 TAGGCCCTTAAT-CTTGCTGTTATTTTCCATGAGCTAAAGGCTGGTCACTGAGCTAC 1833  
DB 1621 TAGGCCCTTAATCTTGCTGTTATTTTCCATGAGCTAAAGGCTGGTCACTGAGCTAC 1680  
QY 1834 GCTGGCTCACAGCTCTAGGGCCCTGCTCCTTAACCTACAGTGGGTTTTGTGAGGCTC 1893

DB 1681 GTGGCTCACAGCTTTAGGGCCCTGCTCCTCTAACTCACAGTGGTTTTGTAGGCTC 1740  
QY 1894 TGTGCCCCAGAGCAGACCTGCATATCTGAGCAAAAATAGCAAAAGCCTCTCTCAGCCAC 1953  
DB 1741 TGTGCCCCAGAGCAGACCTGCATATCTGAGCAAAAATAGCAAAAGCCTCTCTCAGCCAC 1800  
QY 1954 TGGCCTGAATCTACACTGGAAGCAACTTGTCTGGGACCCCGCTCCCCAACCCCTTTTGC 2013  
DB 1801 TGGCCTGAATCTACACTGGAAGCAACTTGTCTGGGACCCCGCTCCCCAACCCCTTTTGC 1860  
QY 2014 CTGGGTAGGAGGCTAAGATCACCCCTAAATTTTACTCATCTCTAGTGTGCCTCACA 2073  
DB 1861 CTGGGTAGGAGGCTAAGATCACCCCTAAATTTTACTCATCTCTAGTGTGCCTCACA 1920  
QY 2074 TTGGGCTCAGCAGCTCCCCAGCACCAATTACAGGTCAACCCCTCTCTTTTGCACGTG 2133  
DB 1921 TTGGGCTCAGCAGCTCCCCAGCACCAATTACAGGTCAACCCCTCTCTTTTGCACGTG 1980  
QY 2134 CCCAACTTGTGTCAAATTCGAGATCTAAATCTCCCCCTACGCTTCGCCAGGAATCTTT 2193  
DB 1981 CCCAACTTGTGTCAAATTCGAGATCTAAATCTCCCCCTACGCTTCGCCAGGAATCTTT 2040  
QY 2194 CAGACCTCACTAGCACAGCCCGTTGCTCTTGTCAAGAGAAATTTGTAGATCATCTCA 2253  
DB 2041 CAGACCTCACTAGCACAGCCCGTTGCTCTTGTCAAGAGAAATTTGTAGATCATCTCA 2100  
QY 2254 CTTCAAATTCCTGGGGCTGATATCTCTCTCATCTTGTGACCCCAACCTCTGTAATAGAT 2313  
DB 2101 CTTCAAATTCCTGGGGCTGATATCTCTCTCATCTTGTGACCCCAACCTCTGTAATAGAT 2160  
QY 2314 TACCGCATTTACGGCTGCATTTCTGTAAGTGGGATGCTCTCTTAATGGAGGAGTTCAT 2373  
DB 2161 TACCGCATTTACGGCTGCATTTCTGTAAGTGGGATGCTCTCTTAATGGAGGAGTTCAT 2220  
QY 2374 TGTATAATAAGTTATTTCACCTGAGTATGCAATAAAGATGTGGTGGCACTCTTTTCATGGT 2433  
DB 2221 TGTATAATAAGTTATTTCACCTGAGTATGCAATAAAGATGTGGTGGCACTCTTTTCATGGT 2280  
QY 2434 GTGGCAGCA 2443  
DB 2281 GTGGCAGCA 2290

RESULT 13  
AF506289  
LOCUS  
DEFINITION Homo sapiens orphan G protein-coupling receptor PEIG-1 mRNA, linear PRI 15-JUL-2002  
complete cds.  
ACCESSION AF506289  
VERSION AF506289.1 GI:21779962  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
1 (bases 1682 to 2040)  
Cafferata,E.G., Gonzalez-Guerrico,A.M., Pivetta,O.H. and  
Santa-Coloma,T.A.  
Identification by differential display of a mRNA specifically  
induced by 12-O-tetradecanoylphorbol-13-acetate (TPA) in T84 human  
colon carcinoma cells  
Cell. Mol. Biol. 42 (5), 797-804 (1996)  
8832110  
2 (bases 1 to 2297)  
Cafferata,E.G., Gonzalez-Guerrico,A.M., Costanzo,R., Pivetta,O.H.  
and Santa-Coloma,T.A.  
Direct Submission  
Submitted (24-APR-2002) Laboratorio de Biologia Celular y  
Molecular, Instituto de Investigaciones Bioquimicas Fundacion  
Campomar, Patricia Argentina 435, Buenos Aires 1405, Argentina  
Location/Qualifiers



QY	1887	GAGGCTCTGTGGCCAGACAGACCTGCATATCTGAGCAAAAATAGCAAAAGCTCTCTC	1946		
Db	1741	GAGGCTCTGTGGCCAGACAGACCTGCATATCTGAGCAAAAATAGCAAAAGCTCTCTC	1800		
QY	1947	AGCCCACTGGCTGAATCTACACTGGAAGCAACTGCTGGCACCCTCCCAACCC	2006		
Db	1801	AGCCCACTGGCTGAATCTACACTGGAAGCAACTGCTGGCACCCTCCCAACCC	1860		
QY	2007	TTCTTGGCTGGGTAGGAGGGTAAAGATCACCCCTAAATTTACTCATCTCTAGTGCTG	2066		
Db	1861	TTCTTGGCTGGGTAGGATAGGCTAAAGATCACCCCTAAATTTACTCATCTCTAGTGCTG	1920		
QY	2067	CCTCATTGGGCTCAGCAGCTCCCAAGCAATTCACAGTCAACCCCTCTCTCTTG	2126		
Db	1921	CCTCATTGGGCTCAGCAGCTCCCAAGCAATTCACAGTCAACCCCTCTCTCTTG	1980		
QY	2127	CACTGTCCCAACTGCTGTCAATTCGGAGATCTAATCTCCCTCAGCTCTGCCAGGA	2186		
Db	1981	CACTGTCCCAACTGCTGTCAATTCGGAGATCTAATCTCCCTCAGCTCTGCCAGGA	2040		
QY	2187	ATTCTTTAGACTCTACTAGCAAGCCGGTGTCTCTTGTGAGAGAAATTTAGATC	2246		
Db	2041	ATTCTTTAGACTCTACTAGCAAGCCGGTGTCTCTTGTGAGAGAAATTTAGATC	2100		
QY	2247	ATTCTCACTCAAAATTCCTGGGCTGATCTCTCTCATCTTGACCCCAACTCTGTAA	2306		
Db	2101	ATTCTCACTCAAAATTCCTGGGCTGATCTCTCTCATCTTGACCCCAACTCTGTAA	2160		
QY	2307	ATAGATTACCGATTACGGCTGCATCTCTGTAAGTGGCGATGGTCTCTAATGGAGAG	2366		
Db	2161	ATAGATTACCGATTACGGCTGCATCTCTGTAAGTGGCGATGGTCTCTAATGGAGAG	2220		
QY	2367	TGTTCACTTAATAAAGTTATTTCACCTCAGATATGCAATAAAGATGTGGTGGCACTC	2424		
Db	2221	TGTTCACTTAATAAAGTTATTTCACCTCAGATATGCAATAAAGATGTGGTGGCACTC	2278		
RESULT 14	BC003665	2296 bp	mRNA	linear	PRI 24-NOV-2004
LOCUS	BC003665				
DEFINITION	Homo sapiens G protein-coupled receptor, family C, group 5, member A, mRNA (cDNA clone MGC:923 IMAGE:2988011), complete cds.				
ACCESSION	BC003665				
VERSION	BC003665.2	GI:33872669			
KEYWORDS	MGC.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	1 (bases 1 to 2296)				
AUTHORS	Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, P.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Rahe, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickinson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalusz, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.				
	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences				
TITLE	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)				
JOURNAL					
PUBMED	12477932				
REFERENCE	2 (bases 1 to 2296)				
AUTHORS	Director MGC Project.				
TITLE	Direct Submission				
JOURNAL	Submitted (26-FEB-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA				
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov				
COMMENT	On Aug 19, 2003 this sequence version replaced gi:13177795. Contact: MGC help desk Email: cgapbs-r@mail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Rubin Laboratory DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA sequencing by: Institute for Systems Biology http://www.systemsbio.org contact: amadan@systemsbiology.org Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting				
FEATURES	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Series: IRAL Plate: 3 Row: 1 Column: 11 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 12056470. Location/Qualifiers 1..2296 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="MGC:923 IMAGE:2988011" /tissue_type="Colon, adenocarcinoma" /clone_lib="NIH MGC 15" /lab_host="DH10B-R" /note="Vector: pOTB7" 1..2296 /gene="GPCR5A" /note="synonyms: RAIG1, GPCR5A, RA13" /db_xref="GeneID:9052" /db_xref="MIM:604138" 95..1168 /gene="GPCR5A" /codon_start=1 /product="G protein-coupled receptor, family C, group 5, member A" /protein_id="AAH03665.1" /db_xref="GI:13177796" /db_xref="GeneID:9052" /db_xref="MIM:604138" /translation="MATTVPDGRNGLSKYYRLCKAEAWGLVETVATAGVVTVA FMLTFLVCKVQDSNRKMLPTQLFLGLVGLTFLGAFIIGLDGSGTPTFFLFG ILFSICFCLLAHVSLLKLVGRKPLSLVLILGLAVGSLVQDVIAIYIVLTMRT NVNVSLSAPRNEDFVLLLTLYLFLMALTFMGSTFCGFTGWRKHAHLYLTWL LSIAIWAHWTLLMLPDPDRMDTILSSALAANGWVLLAYVSPFWLLTKQNPMD YPVEDAFCKPQLVKKSYGVENRAYSQEETIQGFEETGDTLYAPYSTHFLQNPQKE FSIPRAHAWPSPYKDYEVKKEGS"				
gene					
CDS					
Query Match	92.3%	Score	2266.4	DB 5	Length 2296
Best Local Similarity	99.5%	Pred. No.	0		
Matches 2284	Conservative	0	Mismatches	11	Indels 1
Gaps	1				
QY	160	GTCTCCCCCAGCAGCTGAGGAGCTCGCTCTGCTCTTTCGCGCGGGAGCAGACCAA	219		
Db	1	GTCTCCCCCAGCAGCTGAGGAGCTCGCTCTGCTCTTTCGCGCGGGAGCAGACCAA	60		
QY	220	GTTTCAGGCGCAAGCCCTTGGCACTAGGTCAGAAATGGCTACAAAGTCCCTGATGGTTG	279		
Db	61	GTTTCAGGCGCAAGCCCTTGGCACTAGGTCAGAAATGGCTACAAAGTCCCTGATGGTTG	120		
QY	280	CCGCAATGCCTGAAATCCAAGTACTACAGACTTTGTGATAAGCTTGAAGCTTGGGCGCAT	339		



```
RESULT 15
AKI72760
LOCUS
DEFINITION Homo sapiens cDNA FLJ23921 fis, clone COL02043, highly similar to
Homo sapiens retinoic acid induced 3 (RAI3).
ACCESSION AKI72760
VERSION AKI72760.1 GI:47077732
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
AUTHORS Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,
Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 1826)
Sugano,S. and Suzuki,Y.
Direct Submission
Submitted (22-APR-2004) Sumio Sugano, Institute of Medical Science,
University of Tokyo, Laboratory of Genome Structure, Human Genome
Center; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
(E-mail: fclidna@mail.ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing;
Research Association for Biotechnology; cDNA library construction;
5'- & 3'-end one pass sequencing; Department of Virology and Human
Genome Center, Institute of Medical Science, University of Tokyo
(partly supported by Science and Technology Agency).
FEATURES
source
1..1826
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="COL02043"
/tissue_type="colon"
/clone_lib="COL"
/note="cloning vector pME18SFL3"
ORIGIN
Query Match 73.1%; Score 1795.4; DB 5; Length 1826;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 1805; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
QY 636 CCCTTCCCTGTTGGTGAATCTGGGTCCTGGCGTGGGCTTGAGCCTAGTCCAGGATGTTA 695
DB 3 CCCTTCCCTGTTGGTGAATCTGGGTCCTGGCGTGGGCTTGAGCCTAGTCCAGGATGTTA 62
QY 696 TCGCTATTGAATATATTGTCGACCAATAGGACCAACGCTCAATGCTTTCTGAGC 755
DB 63 TCGCTATTGAATATATTGTCGACCAATAGGACCAACGCTCAATGCTTTCTGAGC 122
QY 756 TTTCGGCTCCTCGTCCGAATGAAGACTTTGTCCTCTGCTCCTACCTACGTCCTCTTCTTGA 815
DB 123 TTTCGGCTCCTCGTCCGAATGAAGACTTTGTCCTCTGCTCCTACCTACGTCCTCTTCTTGA 182
QY 816 TGGCGTGACCTTCCTCATGTCTCTCTCTCCTTCACTTCTGTTGGTTCCTTCAAGGCTGAAGA 875
DB 183 TGGCGTGACCTTCCTCATGTCTCTCTCTTCACTTCTGTTGGTTCCTTCAAGGCTGAAGA 242
QY 876 GACATGGGGCCACATCTACCTCAGATGCTCTCTCCATTGCCATCTGGGTGGCTGGA 935
DB 243 GACATGGGGCCACATCTACCTCAGATGCTCTCTCTCCATTGCCATCTGGGTGGCTGGA 302
QX 936 TCACCTCGTCTATGCTTCTCTGACTTTTGACCGCAGGTGGGATGACACCATCTCAGCTCCG 995
DB 303 TCACCTCGTCTATGCTTCTCTGACTTTTGACCGCAGGTGGGATGACACCATCTCAGCTCCG 362
```

Db	1443	GGGCTCAGCAGCTCCCGCAGCACCATTTCACAGGTCAACCCCTCTCTTTGCACTGTCCC	1502
Qy	2136	CAAACTTGCTGTCGAATTCGAGATCTTAATCTCCCTCTAGCTCTGCCAGGAATTCCTTCA	2195
Db	1503	CAAACTTGCTGTCGAATTCGAGATCTTAATCTCCCTCTAGCTCTGCCAGGAATTCCTTCA	1562
Qy	2196	GACCTCACTAGCACAGCCGGTTGCTTCTGTGTCAGGAGAAATTTGTAGATCATTTCTCACT	2255
Db	1563	GACCTCACTAGCACAGCCGGTTGCTTCTGTGTCAGGAGAAATTTGTAGATCATTTCTCACT	1622
Qy	2256	TCAAATTCCTGGGCTGATACTTCTCTCATCTTTGCACCCCAACCTCTGTAAATAGATTTA	2315
Db	1623	TCAAATTCCTGGGCTGATACTTCTCTCATCTTTGCACCCCAACCTCTGTAAATAGATTTA	1682
Qy	2316	CCGCATTACGGCTGCATTCTGTAAGTGGCGATGGTCTCCTAATGGAGGAGTGTTCATTG	2375
Db	1683	CCGCATTACGGCTGCATTCTGTAAGTGGCGATGGTCTCCTAATGGAGGAGTGTTCATTG	1742
Qy	2376	TATAATAAGTTATTACCTGAGTATGCAATAAAGATGTGGTGGCCACTCTTTTCATGGTGG	2435
Db	1743	TATAATAAGTTATTACCTGAGTATGCAATAAAGATGTGGTGGCCACTCTTTAAAAAAA	1802
Qy	2436	TGGCAGCAAAAAA	2456
Db	1803	AAAAAAAAAAAAAAAA	1823

Search completed: June 10, 2006, 18:53:49  
Job time : 13886 secs

**THIS PAGE IS BLANK**





RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
 RA Ohtani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,  
 RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,  
 RA Togashi T., Oyama K., Hata H., Watanabe M., Komatsu T.,  
 RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,  
 RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,  
 RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;  
 RT "Complete sequencing and characterization of 21,243 full-length human  
 RL cDNAs.";  
 RL Nat. Genet. 36:40-45 (2004).  
 RN [4]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE MENA].  
 RC TISSUE:Colon;  
 RX MEDLINE=23388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.P., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Boeak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.W., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettner M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 RL [5]  
 RP TISSUE SPECIFICITY  
 RX MEDLINE=20247251; PubMed=10783259; DOI=10.1006/geno.2000.6164;  
 RA Braeuner-Oberne H., Krosgaard-Larsen P.;  
 RT "Sequence and expression pattern of a novel human orphan G-protein-  
 RT coupled receptor, GPRC5B, a family C receptor with a short amino-  
 RT terminal domain.";  
 RL Genomics 65:121-128 (2000).  
 RN [6]  
 RP TISSUE SPECIFICITY  
 RX PubMed=10945465; DOI=10.1006/geno.2000.6226;  
 RA Robbins M.J., Michalovich D., Hill J., Calver A.R., Medhurst A.D.,  
 RA Gloger I., Sims M.A., Middlemiss D.N., Pangalos M.N.;  
 RT "Molecular cloning and characterization of two novel retinoic acid-  
 RT inducible orphan G-protein-coupled receptors (GPRC5B and GPRC5C).";  
 RL Genomics 67:8-18 (2000).  
 CC -!- FUNCTION: Unknown. This G-protein coupled receptor could be  
 CC involved in modulating differentiation and maintaining homeostasis  
 CC of epithelial cells. The comparable expression level in fetal lung  
 CC and kidney with adult tissues suggests a possible role in  
 CC embryonic development and maturation of these organs. This  
 CC retinoic acid-inducible GPCR provide evidence for a possible  
 CC interaction between retinoid and G-protein signaling pathways.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Localized in the  
 CC plasma membrane and perinuclear vesicles.  
 CC -!- TISSUE SPECIFICITY: Expressed at high level in fetal and adult  
 CC lung tissues. Constitutively expressed in fetal kidney and adult  
 CC placenta, kidney, prostate, testis, ovary, small intestine, colon,  
 CC stomach, and spinal chord at low to moderate levels. Not  
 CC detectable in fetal heart, brain, and liver and adult heart,  
 CC brain, liver, skeletal muscle, pancreas, spleen, thymus, and  
 CC peripheral leukocytes. According to Ref.5: expressed at low but  
 CC detectable level in pancreas and heart.  
 CC -!- INDUCTION: By all-trans retinoic acid (ATRA).  
 CC -!- SIMILARITY: Belongs to the G-protein coupled receptor 3 family.  
 CC  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----

DR EMBL; AF095448; AAC98506.1; -; mRNA.  
 DR EMBL; AF506289; AA077594.1; -; mRNA.  
 DR EMBL; AK001761; BAA91890.1; -; mRNA.  
 DR EMBL; BC003665; AA03665.1; -; mRNA.  
 DR Ensembl; ENSG00000013588; Homo sapiens.  
 DR HGNC; HGNC:9836; GPRC5A.  
 DR MIM; 604138; gene.  
 DR GO; GO:0005887; C: integral to plasma membrane; TAS.  
 DR GO; GO:0004930; F: G-protein coupled receptor activity; TAS.  
 DR GO; GO:0007165; P: signal transduction; TAS.  
 DR InterPro; IPR000337; GPCR\_Mgr.  
 DR Pfam; PF00003; 7tm\_3; 1.  
 DR PROSITE; PS00979; G\_PROTEIN\_RECF\_F3\_1; FALSE\_NEG.  
 DR PROSITE; PS00980; G\_PROTEIN\_RECF\_F3\_2; FALSE\_NEG.  
 DR PROSITE; PS00981; G\_PROTEIN\_RECF\_F3\_3; FALSE\_NEG.  
 DR PROSITE; PS02059; G\_PROTEIN\_RECF\_F3\_4; FALSE\_NEG.  
 KW G-protein coupled receptor; Glycoprotein; Membrane; Polymorphism;  
 KW Receptor; Transducer; Transmembrane.  
 FT CHAIN 1 357 Retinoic acid-induced protein 3.  
 FT /FTID=PRO\_0000206895.  
 FT Extracellular (Potential).  
 FT TRANSMEM 34 54 1 (Potential).  
 FT TOPO\_DOM 55 68 Cytoplasmic (Potential).  
 FT TRANSMEM 69 89 2 (Potential).  
 FT TOPO\_DOM 90 97 Extracellular (Potential).  
 FT TRANSMEM 98 118 3 (Potential).  
 FT TOPO\_DOM 119 129 Cytoplasmic (Potential).  
 FT TRANSMEM 130 150 4 (Potential).  
 FT TOPO\_DOM 151 176 Extracellular (Potential).  
 FT TRANSMEM 177 197 5 (Potential).  
 FT TOPO\_DOM 198 212 Cytoplasmic (Potential).  
 FT TRANSMEM 213 233 6 (Potential).  
 FT TOPO\_DOM 234 247 Extracellular (Potential).  
 FT TRANSMEM 248 268 7 (Potential).  
 FT TOPO\_DOM 269 357 Cytoplasmic (Potential).  
 FT CARBOHYD 158 158 N-linked (GlcNAc...) (Potential).  
 FT VARIANT 118 118 S->G (in dbSNP:850932).  
 FT /FTID=VAR\_018296.  
 FT CONFLICT 292 292 S->G (in Ref. 2).  
 FT SEQUENCE 357 AA; 40251 MW; 7BB524BF6F307E5 CRC64;  
 Query Match 100.0%; Score 1865; DB 1; Length 357;  
 Best Local Similarity 100.0%; Pred. No. 9.2e-138; Indels 0; Gaps 0;  
 Matches 357; Conservative 0; Mismatches 0;  
 QY 1 MATTVPDGCRCNGLSKYYRLCDKAEAWGIVLETATAGVWTSVAFMLTLPILVCKVQDSN 60  
 Db |||||  
 QY 1 MATTVPDGCRCNGLSKYYRLCDKAEAWGIVLETATAGVWTSVAFMLTLPILVCKVQDSN 60  
 Db |||||  
 QY 61 RRKMLPTQFLFLGLVIGIFGLTFAPFIIGLDGSTGTRFRLFGILFSICFSCLLAHAVSLT 120  
 Db |||||  
 QY 61 RRKMLPTQFLFLGLVIGIFGLTFAPFIIGLDGSTGTRFRLFGILFSICFSCLLAHAVSLT 120  
 Db |||||  
 QY 121 KLVGRKPLSLVILGLAVGFSLVQDVIAIEVILTMRTNNVNFSELSAPRNEDFVLL 180  
 Db |||||  
 QY 121 KLVGRKPLSLVILGLAVGFSLVQDVIAIEVILTMRTNNVNFSELSAPRNEDFVLL 180  
 Db |||||  
 QY 181 LTYVLFMALTFMSSFTFCGSFTGCKRKGHAHYLTMLLSIAIWAIVTLLMLPDFDRW 240  
 Db |||||  
 QY 181 LTYVLFMALTFMSSFTFCGSFTGCKRKGHAHYLTMLLSIAIWAIVTLLMLPDFDRW 240  
 Db |||||  
 QY 241 DDTILSSALAANGWVFLAYSPFWLLTKORNPMDYPVEDAFCKPOLVKKSYGVENRAY 300  
 Db |||||  
 QY 241 DDTILSSALAANGWVFLAYSPFWLLTKORNPMDYPVEDAFCKPOLVKKSYGVENRAY 300  
 Db |||||  
 QY 301 SQEETIQCFEETGDTLYAPYSTHFOQNOPPOKFSIPRAHAWPSFYKDYEVKKEGS 357  
 Db |||||  
 QY 301 SQEETIQCFEETGDTLYAPYSTHFOQNOPPOKFSIPRAHAWPSFYKDYEVKKEGS 357  
 Db |||||  
 RESULT 2  
 Q3MHQ6\_BOVIN PRELIMINARY; PRT; 357 AA.  
 ID Q3MHQ6\_BOVIN

```
AC Q3MHO6;
DT 25-OCT-2005, integrated into UniProtKB/TrEMBL.
DT 25-OCT-2005, sequence version 1.
DT 07-MAR-2006, entry version 6.
DE Hypothetical protein MGCI27231.
GN Name=MGCI27231;
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
OC Pecora; Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Crossbred x Angus; TISSUE=ileum;
RA Moore S., Alexander L., Brownstein M., Guan L., Lobo S., Meng Y.,
RA Tanguchi M., Wang Z., Yu J., Frange C., Schreiber K., Shenmen C.,
RA Wagner L., Balu M., Barbazuk S., Barber S., Babakiaff R., Beland J.,
RA Chun E., Del Rio L., Gibson S., Hanson R., Kirkpatrick R., Liu J.,
RA Mateuo C., Mayo M., Santos R.R., Scott J., Tsai M., Wong D.,
RA Siddiqui A., Holt R., Jones S.J., Marra M.A.;
RL Submitted (SEP-2005) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; BC105147; RA105148.1; -; mRNA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008067; F:metabotropic glutamate, GABA-B-like recepto. .; IEA.
DR InterPro; IPR000337; GPCR_Mgr.
DR Pfam; PF00003; 7tm 3; 1.
KW Hypothetical protein.
SQ SEQUENCE 357 AA; 39917 MW; CB5C6B59CFEF5B6C CRC64;

Query Match 78.3%; Score 1460; DB 2; Length 357;
Best Local Similarity 76.5%; Pred. No. 5.1e-106;
Matches 273; Conservative 38; Mismatches 46; Indels 0; Gaps 0;

QY 1 MATTPDGRNGLSKRYRLCDKAEAWGVLETVATAGVTVTSVAFMLTPIILVCKVQDSN 60
Db 1 MSTVAPAGCRSLDSDRYRLCDKAAWGVLETVATAGVTVTSVAFMLTPIILVCKVQDSN 60

QY 61 RRKMLPTQFLFLGLVGLFGLTFAFIIGLDGSGTPTTRFELFGLFSLFCFSCLLAHAVSLT 120
Db 61 KRKLLPTQFLFLGLVGLFGLTFAFIITLDGGTPTTRFELFGLFSCLLAHAVSLT 120

QY 121 KLVGRKPLSLVILGLAVGSLVQDVIAIEYIVLTMTNRTNVVSELSAPRNEDFVLL 180
Db 121 KLVGRQPLSLVILGLAVGSLVQDVIAIEYVLTMTNRTNVVSELSAPRNEDFVML 180

QY 181 LTYVLFMLALTFMSSFTFCGSGFTGKRGHGHYLTMLLSIAIWAWITLLMLPDFDRW 240
Db 181 LIYVLFMLALTFELTASFQSGFTGKRGHGHYLTMLLSIAIWAWITLLVPTLSPQW 240

QY 241 DDTILSSALAANGWFLLAYVSPFEWLLTKQRNPMDPYVEDAFCKPQLVKKSYGVENRAY 300
Db 241 DDTILSSALVANGWFLLAYIAPEFHLITRQQNPMDPYVEDAFQCPQFMKQSYGVNRAY 300

QY 301 SQEITIQGTEETGDTLYAPYSTHFQIQNQPPQKESIPRAHAWPSYKDYVKKSGS 357
Db 301 SQEIIQGTETGDTLYAPYSTHFQIQNQDSAKDFSIPRVQTRVSPYSYDEGRKQVS 357

RESULT 3
RA13 MOUSE
AC Q8BH14; Q8K1G5;
DT 13-APR-2004, integrated into UniProtKB/Swiss-Prot.
DT 01-MAR-2003, sequence version 1.
DT 07-FEB-2006, entry version 26.
DE Retinoic acid-induced protein 3 (G-protein coupled receptor family C
DE group 5 member A) (Retinoic acid-induced gene 1 protein) (RAIG-1).
GN Name=Gprc5a; Synonyms=Ra13, Raigl;
OS Mus musculus (Mouse).
```

```
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridea; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=129/SV; TISSUE=Embryonic stem cell;
RX PubMed=14706456; DOI=10.1016/S0888-7543(03)00237-4;
RA Tao Q., Cheng Y., Clifford J., Lotan R.;
RT "Characterization of the murine orphan G-protein-coupled receptor gene
RL Ra13 and its regulation by retinoic acid.";
RN Genomics 83:270-280(2004).
RN [2]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RC STRAIN=PVB/N; TISSUE=Colon;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.D., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.P., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butcherfield Y.S.N., Krzywinski M.I., Skalska U., Smalilus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RL and mouse cDNA sequences.";
CC -!- FUNCTION: Unknown. This G-protein coupled receptor could be
CC involved in modulating differentiation and maintaining homeostasis
CC of epithelial cells. This retinoic acid-inducible GPCR provide
CC evidence for a possible interaction between retinoid and G-protein
CC signaling pathways.
CC -!- SUBCELLULAR LOCATION: Membrane; multi-pass membrane protein.
CC -!- TISSUE SPECIFICITY: Expressed predominantly in normal fetal and
CC adult lung. Almost undetectable or expressed at very low levels in
CC other tissues.
CC -!- INDUCTION: By all-trans retinoic acid (ATRA).
CC -!- SIMILARITY: Belongs to the G-protein coupled receptor 3 family.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AY079518; AAL87526.1; ALT_INIT; Genomic DNA.
DR EMBL; AY079517; AAL87526.1; JOINED; Genomic DNA.
DR EMBL; BC034063; AAH34063.1; -; mRNA.
DR EMBL; BC036173; AAH36173.1; -; mRNA.
DR EMBL; BC039217; AAH39217.1; -; mRNA.
DR Ensemble; ENSMUSG00000046733; Mus musculus.
DR MGI; MGI:1891250; Gprc5a.
DR InterPro; IPR000337; GPCR_Mgr.
DR Pfam; PF00003; 7tm 3; 1.
DR PROSITE; PS00979; G_PROTEIN_RECEP_F3_1; FALSE NEG.
DR PROSITE; PS00980; G_PROTEIN_RECEP_F3_2; FALSE NEG.
DR PROSITE; PS00981; G_PROTEIN_RECEP_F3_3; FALSE NEG.
DR PROSITE; PS00981; G_PROTEIN_RECEP_F3_4; FALSE NEG.
DR PROSITE; PS02559; G_PROTEIN_RECEP_F3_4; FALSE NEG.
KW G-protein coupled receptor; Glycoprotein; Membrane; Receptor;
KW Transducer; Transmembrane.
CHAIN 1 356 /FTid=PRO_000206896.
TOPO_DOM 1 35 Extracellular (Potential).
TRANSMEM 36 56 1 (Potential).
TOPO_DOM 57 68 Cytoplasmic (Potential).
TRANSMEM 69 89 2 (Potential).
TOPO_DOM 90 101 Extracellular (Potential).
```

```
FT TRANSMEM 102 122 3 (Potential).
FT TOPO DOM 123 131 Cytoplasmic (Potential).
FT TRANSMEM 132 152 4 (Potential).
FT TOPO DOM 153 178 Extracellular (Potential).
FT TRANSMEM 179 199 5 (Potential).
FT TOPO DOM 200 214 Cytoplasmic (Potential).
FT TRANSMEM 215 235 6 (Potential).
FT TOPO DOM 236 244 Extracellular (Potential).
FT TRANSMEM 245 265 7 (Potential).
FT TOPO DOM 266 356 Cytoplasmic (Potential).
FT CARBOHYD 160 160 N-linked (GlcNAc...) (Potential).
FT CONFLICT 196 196 T -> A (in Ref. 1).
SQ SEQUENCE 356 AA; 40101 MW; 08B632F228381C01 CRC64;

Query Match 75.2%; Score 1402.5; DB 1; Length 356;
Best Local Similarity 76.1%; Pred. No. 1.6e-101;
Matches 270; Conservative 26; Mismatches 56; Indels 3; Gaps 2;

QY 3 TTVPDGRNGLSKYRILCDKAEANGIVLETATAGVTSVAFMLTLPILVCKVQDSNRR 62
DB 5 TTAPSGCRSDLSRYHRLCDLAEWGIALETTLAAVGAVATVACMFALVFLICKVQDSNKR 64

QY 63 KMLPTQFLGLVGLGFTAFIIGLGGSTGPTFFELFGILFSCFCLLAHAYSLTKL 122
DB 65 KMLPAQFLGLVGLGFTAFIIGLGGSTGPTFFELFGILFSCFCLLAHAYSLTKL 124

QY 123 VGRGRPLSLVILGLAVGFSLVQDVIAIEYIVLTWNRTNVNVFSELSAPRNEEDFVLLIT 182
DB 125 VGRGRPLSLVILGLAVGFSLVQDVIAIEYIVLTWNRTNVNVFSELSAPRNEEDFVLLIT 184

QY 183 YVFLMALTFMLMSFTFCGSGTGWKRHGAHYLTMLLSIAIWMITLMLPDFRRWDD 242
DB 185 YVFLMLVLTFTFSLVFCGSGTGWKRHGAHYLTMLLSIAIWMITLMLPDFRRWDD 244

QY 243 TILSSALANGVFLAYVSPFWLLTKORNPMDYPVEDAFCKPOLVKSGYVENRAYSQ 302
DB 245 TILSTALVANGVFLAYVSPFWLLTKORNPMDYPVEDAFCKPOLVKSGYVENRAYSQ 304

QY 303 BEITQGFETGDTLYAPYSTHFLQNPQPKQFSPRAHAWPSPYKDYEVKKEGS 357
DB 305 BEITQGL-EMGDTLYAPYSTHFLQNH--QKDFSIPRAQAPASPYNDYEGRKGS 356
```

```
RESULT 4
QSRK14 RAT PRELIMINARY; PRT; 188 AA.
AC QSRK14
DT 21-DEC-2004, integrated into UniProtKB/TrEMBL.
DT 07-FEB-2006, sequence version 1.
DE Ra13 protein (Fragment).
GN Name=Ra13;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Lung;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
```

```
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalilus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Lung;
RG NIH MGC Project;
RL Submitted (NOV-2004) to the EMBL/GenBank/DBJ databases.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC EMBL; BC086372; AAH86372.1; -; mRNA.
DR Ensembl; ENSRNOG00000008412; Rattus norvegicus.
FT NON TER 1
SQ SEQUENCE 188 AA; 21795 MW; 1F20555A74A6DCB6 CRC64;

Query Match 40.2%; Score 749; DB 2; Length 188;
Best Local Similarity 73.7%; Pred. No. 1.4e-50;
Matches 140; Conservative 17; Mismatches 31; Indels 2; Gaps 1;

QY 168 LSAPRNEEDFVLLTYVLFVLMALTFMSFTFCGSGTGWKRHGAHYLTMLLSIAIWM 227
DB 1 LPAPRNEEDFVLLTYVLFVLMALTFMSFTFCGSGTGWKRHGAHYLTMLLSIAIWM 227

QY 228 ITLLMLPDFRRWDDTILSSALANGVFLAYVSPFWLLTKORNPMDYPVEDAFCKPO 287
DB 61 IILLIPIDIRKWDITLSTVLVANGVFLVYILPEFRQLPRQSRSTDPVEDAFCKPO 120

QY 288 LVKKSQYGVENRAYSOEITQGFETGDTLYAPYSTHFLQNPQPKQFSPRAHAWPSPY 347
DB 121 LMKQSYGVENRAYSOEITQGFETGDTLYAPYSTHFLQNPQPKQFSPRAHAWPSPY 178

QY 348 KDYEYKKEGS 357
DB 179 NDYEGRKGS 188

RESULT 5
Q6PA25 XENLA PRELIMINARY; PRT; 357 AA.
AC Q6PA25
DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.
DT 05-JUL-2004, sequence version 1.
DT 07-FEB-2006, entry version 13.
DE MGC68729 protein.
GN Name=MGC68729;
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Lung;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
```



```
FT TOPO DOM 189 204 Cytoplasmic (Potential).
FT TRANSMEM 205 225 6 (Potential).
FT TOPO DOM 226 239 Extracellular (Potential).
FT TRANSMEM 240 260 7 (Potential).
FT TOPO DOM 261 344 Cytoplasmic (Potential).
FT VARSPLIC 299 300 AR -> EC (in isoform 2).
FT VARSPLIC 301 344 /FTID=VSP 010009.
FT VARSPLIC 344 Missing (in isoform 2).
FT CONFLICT 24 24 /FTID=VSP 010010.
FT CONFLICT 234 235 I -> C (in Ref. 3).
FT CONFLICT 235 OP -> HA (in Ref. 3).
SQ SEQUENCE 344 AA; 39056 MW; CBBC1B6A3F683764 CRC64;

Query Match 39.5%; Score 736.5; DB 1; Length 344;
Best Local Similarity 45.4%; Pred. No. 2.5e-49;
Matches 153; Conservative 57; Mismatches 116; Indels 11; Gaps 6;

QY 17 YRLCDKAEANGIVLETATAGVTVTSVAFMLTLPLVCKVQDSNRKMLPTQFLFLLGVL 76
DB 12 YLFCDSNGPWAIVLESIAVIGVITVILLALLAFLMRKVQDCSQMNVLTQFLFLAVL 71
QY 77 GIFGLTFPIIIGDSTGPTREFFLGILFISCFSCLLAHAVSLTKLVGRKPLSLVILG 136
DB 72 GLFGLTFPIIIGDSTGPTREFFLGILFISCFSCLLAHAVSLTKLVGRKPLSLVILG 131
QY 137 LAVGSLVQDVIAEYIVLTMNRNVNVSFELSAPRNEFVLLTYVFLMALTFMSS 196
DB 132 IAGVSLVQDVIAEYIVLTMNRNVNVSFELSAPRNEFVLLTYVFLMALTFMSS 189
QY 197 FTFCSFTGWEKGAHIVLTMLSTAIWNTLML--PDPDR--NDDTILSALAAN 252
DB 150 ATFCCGPNWQHGRILFATVLVSIIVVWVISMILLRGNPQLQRPDHDDAIVICGLVTN 249
QY 253 GWVFLVAVSPFVLLTKQRNPMYDVEDAFCKPOLVKVSYGVENRAYSQEITQGFEE 312
DB 250 ANVFLIIYIIPSLTLYRSR--QECPTQGNVCQVPVQYSFRMDTQETRADSDGAGB- 307
QY 313 GDTLYAPYSTHQLQNPQKQFESIPRAHAWSPYKD 349
DB 308 -DVALTAYGTPQLQSDAPSRVLYPSATL--SPOQD 341

RESULT 7
ID Q3UUY8_MOUSE PRELIMINARY; PRT; 344 AA.
AC Q3UUY8;
DT 11-OCT-2005, integrated into UniProtKB/TrEMBL.
DT 11-OCT-2005, sequence version 1.
DT 07-FEB-2006, entry version 5.
DE 6 days neonate skin cDNA, RIKEN full-length enriched library,
DE clone:A030001P22 product:G protein-coupled receptor, family C, group
DE 5, member D, full insert sequence.
GN Name=Gprc5d;
OS Mus musculus (Mouse);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OX NCBI TaxID=10090;
RN [1]_
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Skin;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Methods Enzymol. 303:19-44(1999).
RN [2]_
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Skin;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Methods Enzymol. 303:19-44(1999).
RN [3]_
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Skin;
RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kiyosawa H.,
RA Nikaide I., Oeato N., Saito R., Suzuki H., Yamanaka I., Gojobori T.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Girardot S., Guscinich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Knapik D.R., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozaki-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.
```

```
RA Ambesi-Impombato A., Apweiler R., Auraliya R.N., Bailey T.L.,
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,
RA Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,
RA di Bernardo D., Down T., Engstrom P., Fagiolini M., Faulkner G.,
RA Fletcher C.F., Fukushima T., Furum P., Futaki S., Gariboldi M.,
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,
RA Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,
RA Hill D., Humnietz L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,
RA Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelson J., Kitamura H.,
RA Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K.,
RA Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,
RA Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,
RA Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,
RA Mottagui-Tabar S., Mulder N., Nakano N., Nakachi H., Ng P.,
RA Nissson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,
RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavese G., Pesole G.,
RA Petrovsky N., Piazza S., Reed J.C., Reid J.F., Ring B.Z., Ringwald M.,
RA Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,
RA Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,
RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,
RA Sperling S., Stupka E., Sugiuira K., Sultana R., Takenaka Y., Taki K.,
RA Tannoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,
RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,
RA Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,
RA Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,
RA Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,
RA Fukuda S., Kanamori-Kayama M., Suzuki M., Aoki J., Arakawa T.,
RA Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashita N.,
RA Kawashima T., Kojima M., Kondo S., Konno K., Nakano K., Ninomiya N.,
RA Nishio T., Okada M., Plessy C., Shibata K., Shiraki T., Suzuki S.,
RA Tagami M., Waki K., Watanabe A., Okamura-Oho Y., Suzuki H., Kawai J.,
RA Hayashizaki Y.;
RT "The transcriptional landscape of the mammalian genome.";
RL Science 309:1559-1563(2005).
RN [3]_
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Skin;
RX PubMed=16141073; DOI=10.1126/science.1112009;
RG RIKEN Genome Exploration Research Group, and Genome Science Group
RG (Genome Network Core Team) and the FANTOM Consortium;
RT "Antisense Transcription in the Mammalian Transcriptome.";
RL Science 309:1564-1566(2005).
RN [4]_
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Skin;
RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kiyosawa H.,
RA Nikaide I., Oeato N., Saito R., Suzuki H., Yamanaka I., Gojobori T.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Girardot S., Guscinich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Knapik D.R., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozaki-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.
```





RA Altechul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.T., Skalska U., Smailus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
CC -!- FUNCTION: Unknown.  
CC -!- SUBCELLULAR LOCATION: Membrane; multi-pass membrane protein.  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=3;  
CC Name=1;  
CC IsoId=Q9NZD1-1; Sequences=Displayed;  
CC Name=2;  
CC IsoId=Q9NZD1-2; Sequences=VSP\_010006, VSP\_010007;  
CC Name=3;  
CC IsoId=Q9NZD1-3; Sequences=VSP\_010008;  
CC -!- TISSUE SPECIFICITY: Widely expressed in the peripheral system.  
CC Expression pattern is high in pancreas, medium in kidney, small  
CC intestine, spleen and testis, low in lung, colon, leukocyte,  
CC prostate and thymus and not detectable in brain, heart, liver,  
CC placenta, skeletal muscle and ovary.  
CC -!- SIMILARITY: Belongs to the G-protein coupled receptor 3 family.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NonDerivs License  
CC -----  
DR EMBL; AF209923; AAF72873.1; -; mRNA.  
DR EMBL; AB099817; BAC79169.1; -; mRNA.  
DR EMBL; AB083629; BAB89342.1; -; Genomic\_DNA.  
DR EMBL; BC089341; AAB69341.1; -; mRNA.  
DR Ensembl; ENSG00000111291; Homo sapiens.  
DR HGNC; HGNC:13310; GPRC5D.  
DR MIM; 607437; Gene.  
DR InterPro; IPR002956; Bride of 7less.  
DR InterPro; IPR000337; GPCR\_Mgr.  
DR Pfam; PF00003; 7tm\_3; 1.  
DR PRINTS; PR01223; BRIDE0F7LESS.  
DR PROSITE; PS00979; G\_PROTEIN\_RECEP\_F3\_1; FALSE NEG.  
DR PROSITE; PS00980; G\_PROTEIN\_RECEP\_F3\_2; FALSE NEG.  
DR PROSITE; PS00981; G\_PROTEIN\_RECEP\_F3\_3; FALSE NEG.  
DR PROSITE; PS0259; G\_PROTEIN\_RECEP\_F3\_4; FALSE NEG.  
KW Alternative splicing; G-protein coupled receptor; Membrane;  
KW Polymorphism; Receptor; Transducer; transmembrane.  
FT CHAIN 1 345  
FT G-protein coupled receptor family C group  
FT 5 member D.  
FT /FTID=PRO\_0000206897.  
FT 1 27  
FT TOPO\_DOM Extracellular (Potential).  
FT TRANSMEM 28 48  
FT 1 (Potential).  
FT TOPO\_DOM 49 63  
FT TRANSMEM 64 84  
FT 2 (Potential).  
FT TOPO\_DOM 85 93  
FT TRANSMEM 94 114  
FT 3 (Potential).  
FT TOPO\_DOM 115 123  
FT TRANSMEM 124 143  
FT 4 (Potential).  
FT TOPO\_DOM 145 167  
FT TRANSMEM 168 188  
FT 5 (Potential).  
FT TOPO\_DOM 189 204  
FT TRANSMEM 205 225  
FT 6 (Potential).  
FT TOPO\_DOM 226 239  
FT TRANSMEM 240 260  
FT 7 (Potential).  
FT TOPO\_DOM 261

FT VARSPLIC 299 345 ARDSGAEEDVALTSYGTPIQPTQVDPDTPQECIPQAKLSPPQ  
FT QDAGGV -> GTFLGDSGSREVLLQEKQKNHVG (in  
FT isoform 3).  
FT /FTID=VSP\_010008.  
FT VARSPLIC 299 300 AR -> DC (in isoform 2).  
FT /FTID=VSP\_010006.  
FT VARSPLIC 301 345 Missing (in isoform 2).  
FT /FTID=VSP\_010007.  
FT VARIANT 18 18 A -> D (in dbSNP:3741822).  
FT /FTID=VAR\_018297.  
FT SEQUENCE 345 AA; 38791 MW; 14B09E4C5E9F9B49 CRC64;  
SQ  
Query Match 39.0%; Score 727.5; DB 1; Length 345;  
Best Local Similarity 45.7%; Pred. NO. 1.2e-48;  
Matches 154; Conservative 55; Mismatches 117; Indels 11; Gaps 6;  
QY 17 YRLCDKAEANGIVLETAVATAGVWTSVAFMLTLPILVCKVQDSNRRKMLPTQFLFLGLVL 76  
DB 12 YFLCDABEGPMGIIIESAIIIGVVITLLLAFLFMKRIQCSQWNVLPQLLFLSLVL 71  
QY 77 GIFGLTFAFIIGLDSGTGTRFFLFGILFSICFCLLAHAYSLTKLVGRKPLSLVLILG 136  
DB 72 GLFGLAFAPFIIELNQQTAPRVYFLGVLFALCFCLLAHASNVLKVRGCVSFSWTTILC 131  
QY 137 LAVGSLVQDVIAEYIVLTWNRTNVVPSLSAPRNEDFVLLTYVFLMALFLMSS 196  
DB 132 IAGCSLLQIIATEYVTLIMTRG--MMFVNMTPCQLNVDFVLLVYVFLMALTFVFSK 189  
QY 197 FTEGSGFTGWRKHGAHIYLTMLLSIAIWMVMTLLML--PDFDR--RWDDTILSSALAN 252  
DB 190 ATFCGPCENWQKGLRIFITVLFSLIIWVWISMLLRGNPQFQRPQWDDPVVICALVTN 249  
QY 253 GWFLLAYVSPFEMLLTKQRNPMDPVEDAFCKPOLVKKSYGVENRAYSQEBITQGFET 312  
DB 250 AWFLLLYIVPELCILYRSCR-QECLQGNACPVYAYQHSFQVENQELSRARDSGAE- 307  
QY 313 GDTLYAPYSTHFQIQNPQPKQEFSPRAHAWPSPYKD 349  
DB 308 -DVALTSYGTPIQPTQVDPDTPQECIPQAKL--SPOOD 341  
RESULT 9  
Q3KNV3\_HUMAN  
ID Q3KNV3\_HUMAN PRELIMINARY; PRT; 345 AA.  
AC Q3KNV3;  
DT 08-NOV-2005, integrated into UniProtKB/TrEMBL.  
DT 08-NOV-2005, sequence version 1.  
DT 07-FEB-2006, entry version 4.  
DE G-protein-coupled receptor, family C, group 5, member D.  
GN Name=GPRC5D;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;  
OC Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=PCR rescued clones;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,



RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RL and mouse cDNA sequences.";  
RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
[2]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=Embryo.  
RG NIH MGC Project;  
RL Submitted (Oct-2005) to the EMBL/GenBank/DBJ databases.  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC  
DR EMBL; BC107078; AAI07079.1; -; mRNA.  
DR EMBL; BC107077; AAI07078.1; -; mRNA.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0008067; F:metabotropic glutamate, GABA-B-like recepto. . ; IEA.  
DR GO; GO:0004872; F:receptor activity; IEA.  
DR GO; GO:0005118; F:sevenless binding; IEA.  
DR GO; GO:0007601; P:visual perception; IEA.  
DR InterPro; IPR002956; Bridge\_of\_7less.  
DR InterPro; IPR000337; GPCR\_Mgr.  
DR Pfam; PF00003; 7tm 3; 1.  
DR PRINTS; PR01223; BRIDBOF7LESS.  
KW Receptor.  
SQ SEQUENCE 345 AA; 38791 MW; 14B09E4C5E9F9B49 CRC64;  
  
Query Match 39.0%; Score 727.5; DB 2; Length 345;  
Best Local Similarity 45.7%; Pred. No. 1.2e-48;  
Matches 154; Conservative 55; Mismatches 117; Indels 11; Gaps 6;  
  
QY 17 YRLCDKAEANGIVLETVATAGVTVSVAFMLTLPILVCKVQDSNRRMLPTQPLFLGLVL 76  
DB 12 YFLLCDAECPGWIILSLAIGLVTVVILLAFLLMRKIQDCSQWNVLPQTQLFLSLVL 71  
QY 77 GIFGLTFAFIIGDGTGTPRFLFGILFISCFSCLLAHAVSLTKLVGRKPLSLVILG 136  
DB 72 GLFGLAFATIELNQOTAPVRFLFGLFALCFSCLLAHASNLKLVGRKPLSLVILG 131  
QY 137 LAVGSLVDQVIAIEYIVLTMRNTNVNVSFELSAPRNEDFVLLTYVLFMLALTFMS 196  
DB 132 IAIGCSLLQIIATEYVTLIMTRG--MMFVNMTPCQLNVDFVLLVYVLFMLALTFVSK 189  
QY 197 FTFCSFTGCKRHGAHYITMLLSIAIYVAVITLML--PDPR--RWDITLSSMAAN 252  
DB 190 ATFCGCPENWKQGRILFITVLFSTIIWVWISMLLRGNPQFORQPQWDPPVVCIALVTN 249  
QY 253 GWVFLAYVSPFVLLTKQNPMDYVEDAFCKPQLVKKSYGVENRAYSQEEITQGFET 312  
DB 250 ANVFLLYIVPELCILYRSCR--QECPLQGNACPVTAHQSHFQVENELSRDSDGAE- 307  
QY 313 GDTLXAPYSTHFQIQNPQPKFESIPRAHAWSPYKD 349  
DB 308 -DVALTSYGTPIQPQVDFPQECFIQAKL--SPQD 341  
  
RESULT 10  
QSM7Q6 XENTR PRELIMINARY; PRT; 432 AA.  
AC QSM7Q6;  
DT 01-FEB-2005, integrated into UniProtKB/TREMBL.  
DT 01-FEB-2005, sequence version 1.  
DE 07-FEB-2006, entry version 9.  
DE Hypothetical LOC496814.  
GN Name=LOC496814;  
OS Xenopus tropicalis (Western clawed frog) (Silurana tropicalis).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;  
OC Xenopodinae; Xenopus; Silurana.  
OX NCBI\_TaxID=8364;  
RN [1]

RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=Embryo.  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RL and mouse cDNA sequences.";  
RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
[2]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=Embryo.  
RA Klein S., Gerhard D.S.;  
RL Submitted (DEC-2004) to the EMBL/GenBank/DBJ databases.  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC  
DR EMBL; BC088508; AAH88508.1; -; mRNA.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0008067; F:metabotropic glutamate, GABA-B-like recepto. . ; IEA.  
DR InterPro; IPR000337; GPCR\_Mgr.  
DR Pfam; PF00003; 7tm 3; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 432 AA; 48246 MW; 2DAB4751E7203433 CRC64;  
  
Query Match 32.5%; Score 605.5; DB 2; Length 432;  
Best Local Similarity 37.6%; Pred. No. 5.8e-39;  
Matches 143; Conservative 66; Mismatches 120; Indels 51; Gaps 15;  
  
QY 3 TTVPGCRNGLSKYVRLCDKAEANGIVLETVATAGVTVSVAFMLTLPILVCK---VQDS 59  
DB 30 TAVPGCGAGINSLYYLCDLNAAAGIVLQAITAGIVST--FILTI-VLVAFTPIQNK 86  
QY 60 NRRKMLPTQFLGLVGLFGLTFAFIIGLDGSGTPTFRFLFGILFISCFSCLLAHAVSL 119  
DB 87 KKSLLGTQVFLGLTGLFGLVDFVDMVKDKATCASRRFLGGLLFAICFACLWVHGVS 146  
QY 120 TKLVGRKPLSLVILGLAVGSLVDVIAIEYIVLTMRNTNVNVSFELSAPR-----RN 174  
DB 147 NYLIRNTQPSGWIIFGLAVCLSLVEAIINTEWLIITIVKT-----ANPPADPCLIGN 199  
QY 175 EDFVLLTYVLFMLALTFMLSSFTFCGSGTCKRCAHYITMLLSIAIYVAVITLMLP 234  
DB 200 ADFVMAIYVFMFLIAGFITAPALCGRYGHRKKAIFILLTFLFISIAIYVIMVYVG 259  
QY 235 DFD-----RRWDDTILSSALAANGVFLAYVSPFVLLTK-----QRNMDYPVE--- 280  
DB 260 NVQVGNPLYWDDPTLAIALVSGNAFIPFYIPEISQLTKPTLEQTFESEP--YPIRGVG 317  
QY 281 -DAFCKPQLVKKSYGVENRAYSQEEIT-----QGFE-BTGDTLYAPYSTHFQIQNP 329  
DB 318 YETILKEQSTQRMV-VENKAFSMDPEPPSAKKPVPVSYGVNGTGKSVQYQ--TEMALMNK 374  
QY 330 P-PQKEFS--IPRAHAWSP 346  
DB 375 TLPSYDVIVIPRATANPPP 394

```

RESULT 11
GPC5C_MOUSE
ID Q8K3J9; Q8K0H0; STANDARD; PRT; 440 AA.
DT 13-APR-2004, integrated into UniProtKB/Swiss-Prot.
DT 13-APR-2004, sequence version 2.
DT 07-FEB-2006, entry version 26.
DE G-protein coupled receptor family C group 5 member C precursor
DE (Retinoic acid-induced gene 3 protein) (RAIG-3).
GN Name=Gprc5c; Synonyms=Raig3;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridea; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA].
RC STRAIN=BALB/c; TISSUE=Kidney;
RA Tao Q., Lotan R.;
RT "Molecular cloning and characterization of mouse retinoic acid-
inducible orphan G protein-coupled receptors.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
[2]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RC TISSUE=Kidney;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Klausner R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altchul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettner M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailis D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra W.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
CC -!- FUNCTION: Unknown. This retinoic acid-inducible G-protein coupled
receptor provide evidence for a possible interaction between
retinoid and G-protein signaling pathways (By similarity).
CC -!- SUBCELLULAR LOCATION: Membrane; multi-pass membrane protein.
CC -!- SIMILARITY: Belongs to the G-protein coupled receptor 3 family.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NonDerivs License
CC
DR EMBL; AY079516; AAL87525.1; -; mRNA.
DR EMBL; BC031439; AAL87525.1; -; mRNA.
DR Ensembl; ENSMUSG0000051043; Mus musculus.
DR MGI; MGI:1917605; Gprc5c.
DR GO; GO:0005739; C:mitochondrion; IDA.
DR InterPro; IPR002956; Bridge_of_fless.
DR InterPro; IPR000337; GPCR_Mgr.
DR Pfam; PF00003; 7tm_3; 1.
DR PRINTS; PR01223; BRIDE0FLLESS.
DR PROSITE; PS00979; G_PROTEIN_RECEP_F3_1; FALSE NEG.
DR PROSITE; PS00980; G_PROTEIN_RECEP_F3_2; FALSE NEG.
DR PROSITE; PS00981; G_PROTEIN_RECEP_F3_3; FALSE NEG.
DR PROSITE; PS00981; G_PROTEIN_RECEP_F3_4; 1.
DR PROSITE; PS00981; G_PROTEIN_RECEP_F3_5; 1.
KW G-protein coupled receptor; Glycoprotein; Membrane; Receptor; Signal;
Transducer; Transmembrane.
FT SIGNAL 1 22 Potential.
FT CHAIN 23 440 G-protein coupled receptor family C group
5 member C.

```

```

/FTid=PRO_0000012968.
Extracellular (Potential).
1 (Potential).
Cytoplasmic (Potential).
2 (Potential).
Extracellular (Potential).
3 (Potential).
Cytoplasmic (Potential).
4 (Potential).
Extracellular (Potential).
5 (Potential).
Cytoplasmic (Potential).
6 (Potential).
Extracellular (Potential).
7 (Potential).
Cytoplasmic (Potential).
N-linked (GlcNAc...) (Potential).
S -> L (in Ref. 1).
S 52
SQ SEQUENCE 440 AA; 48422 MW; BA0B13B7631017B4 CRC64;

Query Match 30.8%; Score 574; DB 1; Length 440;
Best Local Similarity 38.3%; Pred. No. 1.7e-36;
Matches 143; Conservative 60; Mismatches 130; Indels 40; Gaps 11;

QY 6 PDGCRGLKSKYRLCDKAEANGIVLETVATAGVTVTSVAFMLTLPLVCK---VQDSNRR 62
DB 27 PPGCSPLDPLYYNLCDRSGAGIVSEAVAGAGIIT--FVLTI--ILVASLFPVQDTKKR 83
QY 63 KMLTQFLPLGLVLGIFGLTFAFIIGLDGSGTPTREFLFGILFSCFSCLLAHAVSLTKL 122
DB 84 SILGTQVFLGLTGLGFLVACVVKPDPSTCASRRFLFGVLFAICFCLVAHVLSLNF 143
QY 123 VRGRKPLSLVLGLAVGFSLVQDVIAIYIVLTMNR-----TNVNVFSELSAP--R 172
DB 144 TRKHGPRGWIFTVALLTLVEVIINTEWLIITLVRGGGQVSLGNSADSTMTSPCAI 203
QY 173 RNEDVFLVLTIVFLMALITFLMSSFTFCGSGTGWKRGHGHVITMLLSIAIVAVITLLM 232
DB 204 ANMDFVMAIIVYMLLLLTAFILGAWPTLCGRPKRWKRGVFLVLLTIVISIAIVWVMT 263
QY 233 LPDFDR---WDDTILSALAANGVFLVAVVSPFVLLTKQNPMDYPVE----- 280
DB 264 YCNEQHSHSPTDDPTLALALANAMTVLVIPEVSQVTKRPSQSYQGDWYPTRGVGY 323
QY 281 DAFCKPQLVKSYGVENRAYSOEE-----ITQGFETGDTLYAPYS--THFQLQNPPQ 332
DB 324 ETILKEQ--TGQSMFVENKAFSMDPASKRPPVSPYSGYNGQLLTSTVYOPTMALMHKGPS 382
QY 333 K---EFSIPRAHA 342
DB 383 EGAYDVILPRATA 395

RESULT 12
Q3KRC4_RAT
ID Q3KRC4_RAT PRELIMINARY; PRT; 457 AA.
AC Q3KRC4;
DT 08-NOV-2005, integrated into UniProtKB/TrEMBL.
DT 08-NOV-2005, sequence version 1.
DT 07-FEB-2006, entry version 4.
DE Gprc5c protein (Fragment).
GN Name=Gprc5c;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridea; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Prostate;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Klausner R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

```

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Brownstein M.J., Uedini T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RL and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=Prostate;  
RG NIH MGC Project;  
RL Submitted (SEP-2005) to the EMBL/GenBank/DBJ databases.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
CC EMBL; BC105781; AAI05782.1; -; mRNA.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0008067; F:metabotropic glutamate, GABA-B-like recepto. .; IEA.  
DR InterPro; IPR000337; GPCR\_Mgr.  
DR Pfam; PF00003; 7tm\_3; 1.  
DR PROSITE; PS50259; G\_PROTEIN\_RECEP\_F3\_4; 1.  
FT NON\_TER 1  
SQ SEQUENCE 457 AA; 49954 MW; CF2E2223ACDA1E2D CRC64;  
  
Query Match 30.2%; Score 564; DB 2; Length 457;  
Best Local Similarity 34.9%; Pred. No. 1.1e-35;  
Matches 145; Conservative 61; Mismatches 138; Indels 72; Gaps 12;  
  
QY 6 PDGCRGLSKYRLCKAEMGIVLETATAGVWTSVAFMLTLPILVCK---VQDSNRR 62  
DB 6 PPGCSPLDPLYYNLCDSRGAMGIVLEAVAGAGIIT--FVLTI-ILVASLPFVDTKR 99  
QY 63 KMLPTQFLGLVGLFGITFPATIGLDGSGTPTREFLGLPFCISFCILAHAVSLTKL 122  
DB 100 SLLGTQVFFLLGLTGLFCLVFACVWKVPDFSTCASRFLFGLVFAICFSLIAHTLSNLF 159  
QY 123 VGRKPLSLVLVLGLAVGFSLVQDVTAIEVILVTMRTNVNVSLSAPRR----- 173  
DB 160 ARKNHGPRGWVFTVALLTLTLEVIINTWLIITLVRG--GQVSTFGNSADWTVTS 215  
QY 174 -----NEDEVLLTYVFLMALTFLLMSSTFCGSGFTGWRKHGHAHYLTLLSIAIWMVI 228  
DB 216 PCAIANMDFWALYYVLLMLLAFLGAWPTLCGRFWRKHGKGVFLVLTATSIATVWVI 275  
QY 229 TLLMLPDFDR---WDDTILSSMALANGVFLAYVSPFWLLTKORNPMDPVE----- 280  
DB 276 VMVYGNKQHSHTDPTDLATLALANAWTFVFVYIPEVSQVTKPSQSQGDMYPTR 335  
QY 281 ----DAFCQPOLVKKSGYVENRAYSOEE-----ITQGFEEGDTLYAPYS--THFQLQN 328  
DB 336 GVGYETILKEQ--TGQSMFVNKAFSMDPEASAKRPVSPYSGYNGQLTSVYQPTENALMH 394  
QY 329 QPPQK---BFSIPRA-----HAMSPYKYDEVKKEGS 357  
DB 395 KGPSEGAYDVLPRATANSQVWGSANSTLRAEDMYVQSHOVATPTKDGKISQDS 450  
  
RESULT 13  
GPC5C HUMAN  
ID GPC5C HUMAN STANDARD; PRT; 441 AA.  
AC Q9NQ84; Q9NZG5;

DT 13-APR-2004, integrated into UniProtKB/Swiss-Prot.  
DT 13-APR-2004, sequence version 2.  
DT 07-FEB-2006, entry version 28.  
DE G-protein coupled receptor family C group 5 member C precursor  
DE (Retinoic acid-induced gene 3 protein) (RAIG-3).  
GN Name=GPC5C; Synonyms=RAIG3; ORFNames=PFSEC0087;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;  
OC Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [MRNA], TISSUE SPECIFICITY, SUBCELLULAR LOCATION,  
RP AND INDUCTION.  
RX PubMed=10945445; DOI=10.1006/geno.2000.6226;  
RA Robbins M.J., Michalovich D., Hill J., Calver A.R., Medhurst A.D.,  
RA Gloger I., Sims M.A., Middlemiss D.N., Pangalos M.N.;  
RT "Molecular cloning and characterization of two novel retinoic acid-  
RT inducible orphan G-protein-coupled receptors (GPC5B and GPC5C).";  
RL Genomics 67:8-18(2000).  
RN [2]  
RP NUCLEOTIDE SEQUENCE [MRNA], TISSUE SPECIFICITY, SUBCELLULAR LOCATION,  
RP AND INDUCTION.  
RX MEDLINE=21210966; PubMed=11311935;  
RA Braeuer-Osborne H., Jensen A.A., Sheppard P.O., Brodin B.,  
RA Krosgaard-Larsen P., O'Hara P.;  
RT "Cloning and characterization of a human orphan family C G-protein  
RT coupled receptor GPC5D.";  
RL Biochim. Biophys. Acta 1518:237-248(2001).  
RN [3]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].  
RC TISSUE=Placenta;  
RA Ota T., Nishikawa T., Suzuki Y., Kawai-Hio Y., Hayashi K., Ishii S.,  
RA Saito K., Yamamoto J., Wakamatsu A., Nagai T., Nakamura Y.,  
RA Nagahara K., Sugano S., Isogai T.;  
RT "HRI human cDNA sequencing project.";  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].  
RA Kalnine N., Chen X., Rolfs A., Halleck A., Hines L., Eisenstein S.,  
RA Koundinya M., Raphael J., Moreira D., Kelley T., LaBaer J., Lin Y.,  
RA Phelan M., Farmer A.;  
RT "Cloning of human full-length cDNAs in BD Creator(TM) system donor  
RT vector.";  
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].  
RC TISSUE=Pancreas;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Dege J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,  
RA Brownstein M.J., Uedini T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RL and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
CC -!- FUNCTION: Unknown. This retinoic acid-inducible G-protein coupled  
CC receptor provide evidence for a possible interaction between  
CC retinoid and G-protein signaling pathways (By similarity).  
CC -!- INTERACTION:





**THIS PAGE IS BLANK**

GenCore version 5.1.1.9  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2006, 15:18:44 ; Search time 1145.2 Seconds  
(without alignments)  
2345.260 Million cell updates/sec

Title: US-10-600-816-30

Perfect score: 42

Sequence: 1 gccacgcttgccgagcccc.....acaaagactatgaagtaaa 42

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 6366136 seqs, 31973710525 residues

Total number of hits satisfying chosen parameters: 12732272

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

Database :

GenEmbl.\*

1: gb\_env.\*

2: gb\_pat.\*

3: gb\_ph.\*

4: gb\_pl.\*

5: gb\_pr.\*

6: gb\_ro.\*

7: gb\_sts.\*

8: gb\_sy.\*

9: gb\_un.\*

10: gb\_vl.\*

11: gb\_ov.\*

12: gb\_htg.\*

13: gb\_in.\*

14: gb\_om.\*

15: gb\_ba.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Query	ID	Description
1	42	100.0	603	2	BD233463
2	42	100.0	1212	2	BD209659
3	42	100.0	1212	2	AR341505
4	42	100.0	1619	2	AX078375
5	42	100.0	1718	2	BD233473
6	42	100.0	1826	5	AK172760
7	42	100.0	2290	2	CQ723177
8	42	100.0	2296	5	BC003665
9	42	100.0	2297	5	AF506289
10	42	100.0	2302	2	AF930411
11	42	100.0	2302	5	AF095448
12	42	100.0	2446	2	BD156680
13	42	100.0	2446	2	AX877483
14	42	100.0	2446	5	AK001761
15	42	100.0	2456	2	CQ981495
16	42	100.0	2456	2	DD210040
17	42	100.0	2456	2	AX549168
18	42	100.0	3057	5	AK122672

19	42	100.0	3371	2	AX188348
20	42	100.0	6730	2	CQ894732
c 21	42	100.0	161577	5	AC007688
22	33.6	80.0	680	2	BD150346
23	33.6	80.0	680	2	AX870284
24	26	61.9	194709	12	AC132640
25	26	61.9	232234	12	AC108634
c 26	26	61.9	240216	12	AC107096
27	24.6	58.6	235468	6	AC105586
28	24.6	58.6	253631	12	AC126155
29	24.4	58.1	189833	12	AC116169
30	24	57.1	229303	12	AC157012
31	23.6	56.2	161566	5	AC008060
c 32	23	54.8	135629	12	CT573208
33	23	54.8	169003	12	CT009616
c 34	23	54.8	185146	6	AC159967
35	23	54.8	230474	6	AC102564
36	22.8	54.3	1500	2	AR630602
37	22.8	54.3	1500	10	AF112297
38	22.8	54.3	1500	10	AF112298
39	22.8	54.3	1500	10	AY243826
40	22.8	54.3	1500	10	AY243843
41	22.8	54.3	1500	10	JEU44974
42	22.8	54.3	1500	10	JEU70389
43	22.8	54.3	1500	10	JEU70413
44	22.8	54.3	1500	10	JEU70419
45	22.8	54.3	1500	10	JEU70420
46	22.8	54.3	1500	10	JEU70420
47	22.8	54.3	1896	10	AY184214
48	22.8	54.3	2069	14	BC105147
49	22.8	54.3	2379	10	S75726
50	22.8	54.3	2436	10	JEU03696
51	22.8	54.3	4320	10	JEU03696
52	22.8	54.3	4512	2	AR003118
53	22.8	54.3	4512	2	AR003211
54	22.8	54.3	4512	10	JEU03696
55	22.8	54.3	4992	2	I08692
56	22.8	54.3	6536	2	BD064747
57	22.8	54.3	7486	2	BD064684
58	22.8	54.3	7486	2	BD064685
59	22.8	54.3	7486	2	BD064746
60	22.8	54.3	10976	10	JEU03696
c 61	22.8	54.3	77422	13	AC025723
c 62	22.8	54.3	141695	6	AC121911
c 63	22.8	54.3	145213	6	AC142410
c 64	22.8	54.3	247690	12	AC095186
65	22.8	54.3	249669	12	AC137204
66	22.8	54.3	298216	12	AC006875
67	22.6	53.8	161575	12	AC141041
c 68	22.6	53.8	173071	12	AC166012
c 69	22.6	53.8	186622	12	AC091441
c 70	22.4	53.3	1857	13	DQ211932
c 71	22.4	53.3	11959	13	AC024754
c 72	22.4	53.3	34483	5	AC135368
c 73	22.4	53.3	110000	15	GR522870
c 74	22.4	53.3	118580	15	PPU344068
c 75	22.4	53.3	136857	5	AC068600
76	22.4	53.3	159453	12	AC006732
77	22.4	53.3	176355	5	AC025518
c 78	22.4	53.3	181914	12	AC024640
c 79	22.4	53.3	186870	12	AC116170
80	22.2	52.9	584	2	AX400400
81	22.2	52.9	69203	12	AC104541
82	22.2	52.9	73878	12	AC011600
83	22.2	52.9	110000	12	AC011600
c 84	22.2	52.9	156766	5	AC079597
c 85	22.2	52.9	205274	12	AC128161
c 86	22.2	52.9	21886	11	AC171018
c 87	22.2	52.9	226291	12	AC096068
88	22.2	52.9	247515	12	AC126713
c 89	22	52.4	167735	6	AC122818
c 90	22	52.4	190256	6	AC121302
c 91	22	52.4	191710	6	AC122489

AX188348	Sequence
CQ894732	Sequence
AC007688	Homo sapi
BD150346	Primer fo
AX870284	Sequence
AC132640	Rattus no
AC108634	Rattus no
AC107096	Rattus no
AC105586	Rattus no
AC126155	Rattus no
AC116169	Homo sapi
AC157012	Bos tauru
AC008060	Homo sapi
CT573208	Danio ter
CT009616	Danio ter
AC159967	Mus muscu
AC102564	Mus muscu
AR630602	Sequence
AF112297	Japanese
AF112298	Japanese
AY243826	Japanese
AY243843	Japanese
AY4974	Japanese en
U70389	Japanese en
U70413	Japanese en
U70419	Japanese en
U70420	Japanese en
Z34097	Japanese en
AY184214	Japanese
BC105147	Bos tauru
S75726	Japanese en
U03696	Japanese en
M16574	Japanese en
AR003118	Sequence
AR003211	Sequence
M73710	Japanese en
I08692	Sequence 2
BD064747	Expressio
BD064684	Preparati
BD064685	Preparati
BD064746	Expressio
L48961	Japanese en
AC025723	Caenorhab
AC121911	Mus muscu
AC142410	Mus muscu
AC095186	Rattus no
AC137204	Rattus no
AC006875	Caenorhab
AC141041	Rattus no
AC166012	Oryctolag
AC091441	Homo sapi
DQ211932	Toxoplasma
AC024754	Caenorhab
AC135368	Homo sapi
Continuation (22 o	
AC1344068	Pseudomon
AC068600	Homo sapi
AC006732	Caenorhab
AC025518	Homo sapi
AC024640	Homo sapi
AC116170	Homo sapi
AX400400	Sequence
AC104541	Mus muscu
Continuation (4 of	
AC011600	3 of
AC079597	Homo sapi
AC128161	Rattus no
AC171018	Gallus ga
AC096068	Rattus no
AC126713	Rattus no
AC122818	Mus muscu
AC121302	Mus muscu
AC122489	Mus muscu

C 92	22	52.4	203589	6	AC084383	AC084383 Mus muscu	165	21.4	51.0	2356	2	CQ574206	CQ574206 Sequence
C 93	22	52.4	209153	12	AC092090	AC092090 Canis fam	166	21.4	51.0	2368	13	AF132165	AF132165 Drosophil
C 94	22	52.4	249159	12	AC111880	AC111880 Rattus no	167	21.4	51.0	3579	10	AY904042	AY904042 Trypanoso
C 95	22	52.4	255568	12	AC127823	AC127823 Rattus no	168	21.4	51.0	4880	13	SEV1GAG	X58484 Srian foam
C 96	21.8	51.9	428	11	AY844358	AY844358 Acris cre	169	21.4	51.0	12972	10	SPVGENOME	X54482 Sman foam
C 97	21.8	51.9	428	11	AY844359	AY844359 Acris gry	170	21.4	51.0	32182	13	DQ011274	DQ011274 Oikopleur
C 98	21.8	51.9	428	11	AY844389	AY844389 Hyia arbo	171	21.4	51.0	53394	5	AC013276	AC013276 Homo sapi
C 99	21.8	51.9	428	11	AY844391	AY844391 Hyia aren	c 172	21.4	51.0	58558	12	AC024305	AC024305 Homo sapi
C 100	21.8	51.9	428	11	AY844420	AY844420 Hyia japo	c 173	21.4	51.0	110000	4	AF008213	AF008213 110 Continuation (111
C 101	21.8	51.9	428	11	AY844434	AY844434 Hyia mixe	c 174	21.4	51.0	119895	15	AP004310	AP004310 Synchocy
C 102	21.8	51.9	428	11	AY844462	AY844462 Hyia squi	c 175	21.4	51.0	120469	12	AC158875	AC158875 Bos tauru
C 103	21.8	51.9	428	11	AY844465	AY844465 Hyia veis	c 176	21.4	51.0	125648	12	AC158857	AC158857 Bos tauru
C 104	21.8	51.9	428	11	AY844466	AY844466 Hyia walk	c 177	21.4	51.0	143668	4	AF004268	AF004268 Oryza sat
C 105	21.8	51.9	428	11	AY844504	AY844504 Pseudacri	c 178	21.4	51.0	145342	12	AC179493	AC179493 Strongylo
C 106	21.8	51.9	428	11	AY844519	AY844519 Scinax fu	c 179	21.4	51.0	147826	12	AC141723	AC141723 Apis mell
C 107	21.8	51.9	734	7	BV630333	BV630333 S21P6119	c 180	21.4	51.0	177982	12	AC155545	AC155545 Zea mays
C 108	21.8	51.9	1339	2	AX646217	AX646217 Sequence	c 181	21.4	51.0	182523	12	AC182208	AC182208 Bos tauru
C 109	21.8	51.9	1339	5	AB065929	AB065929 Homo sapi	c 182	21.4	51.0	193237	12	AC179031	AC179031 Strongylo
C 110	21.8	51.9	1509	11	AY583347	AY583347 Chlogloss	c 183	21.4	51.0	211717	12	AC175161	AC175161 Bos tauru
C 111	21.8	51.9	1512	11	AY571662	AY571662 Hyia meri	c 184	21.4	51.0	223378	12	AC015862	AC015862 Homo sapi
C 112	21.8	51.9	1512	11	AY583339	AY583339 Hyia meri	c 185	21.4	51.0	237488	12	AC095201	AC095201 Rattus no
C 113	21.8	51.9	21670	5	AF508041	AF508041 Homo sapi	c 186	21.4	51.0	243969	11	AC145956	AC145956 Gallus ga
C 114	21.8	51.9	21748	2	CQ595097	CQ595097 Sequence	c 187	21.4	51.0	245036	12	AC094925	AC094925 Rattus no
C 115	21.8	51.9	36628	5	AF037222	AF037222 Human DNA	c 188	21.4	51.0	256110	12	AC095639	AC095639 Rattus no
C 116	21.8	51.9	55704	5	CT009610	CT009610 Human DNA	c 189	21.4	51.0	306349	12	AC145065	AC145065 Pan trogl
C 117	21.8	51.9	95961	12	AC017329	AC017329 Drosophil	c 190	21.2	50.5	286	10	AB029292	AB029292 Japanese
C 118	21.8	51.9	100598	5	AC006271	AC006271 Homo sapi	c 191	21.2	50.5	286	10	AB029293	AB029293 Japanese
C 119	21.8	51.9	143136	13	AC023677	AC023677 Drosophil	c 192	21.2	50.5	286	10	AB029294	AB029294 Japanese
C 120	21.8	51.9	146466	5	AL353637	AL353637 Human DNA	c 193	21.2	50.5	286	10	AB029295	AB029295 Japanese
C 121	21.8	51.9	156415	5	AC011464	AC011464 Homo sapi	c 194	21.2	50.5	286	10	AB029296	AB029296 Japanese
C 122	21.8	51.9	165617	5	CNS0000K	AL049840 Human chr	c 195	21.2	50.5	286	10	AB029300	AB029300 Japanese
C 123	21.8	51.9	167531	5	AL158832	AL158832 Human DNA	c 196	21.2	50.5	286	10	AB029301	AB029301 Japanese
C 124	21.8	51.9	175378	12	AC113179	AC113179 Mus muscu	c 197	21.2	50.5	618	10	AY501273	AY501273 HIV-1 iso
C 125	21.8	51.9	175679	13	AC010025	AC010025 Drosophil	c 198	21.2	50.5	1482	10	AF036915	AF036915 Japanese
C 126	21.8	51.9	178159	12	AC173237	AC173237 Bos tauru	c 199	21.2	50.5	1500	10	AB028248	AB028248 Japanese
C 127	21.8	51.9	181194	6	AL844531	AL844531 Mouse DNA	c 200	21.2	50.5	1500	10	AB028249	AB028249 Japanese
C 128	21.8	51.9	193461	5	AC023403	AC023403 Homo sapi	c 201	21.2	50.5	1500	10	AB028250	AB028250 Japanese
C 129	21.8	51.9	206834	12	AC135066	AC135066 Homo sapi	c 202	21.2	50.5	1500	10	AB028251	AB028251 Japanese
C 130	21.8	51.9	207858	12	AC127132	AC127132 Rattus no	c 203	21.2	50.5	1500	10	AB028271	AB028271 Japanese
C 131	21.8	51.9	229892	12	AC096895	AC096895 Rattus no	c 204	21.2	50.5	1500	10	AY029207	AY029207 Japanese
C 132	21.8	51.9	248669	12	AC127097	AC127097 Rattus no	c 205	21.2	50.5	1500	10	AY243827	AY243827 Japanese
C 133	21.8	51.9	253387	12	AC133379	AC133379 Rattus no	c 206	21.2	50.5	1500	10	AY243831	AY243831 Japanese
C 134	21.8	51.9	297381	13	AE003540	AE003540 Drosophil	c 207	21.2	50.5	1500	10	AY243833	AY243833 Japanese
C 135	21.6	51.4	452	2	BD040609	BD040609 Sequence	c 208	21.2	50.5	1500	10	AY243837	AY243837 Japanese
C 136	21.6	51.4	452	2	AR740870	AR740870 Sequence	c 209	21.2	50.5	1500	10	AY243839	AY243839 Japanese
C 137	21.6	51.4	452	2	AX905076	AX905076 Sequence	c 210	21.2	50.5	1500	10	AY555762	AY555762 Japanese
C 138	21.6	51.4	65851	12	AC172133	AC172133 Bos tauru	c 211	21.2	50.5	1500	10	AY555764	AY555764 Japanese
C 139	21.6	51.4	70000	2	AR215862	AR215862 Sequence	c 212	21.2	50.5	1500	10	AY555765	AY555765 Japanese
C 140	21.6	51.4	73317	5	AY522821	AY522821 Homo sapi	c 213	21.2	50.5	1500	10	AY555766	AY555766 Japanese
C 141	21.6	51.4	95375	5	HS2288A9	AL022322 Human DNA	c 214	21.2	50.5	1500	10	AEU44958	AEU44958 Japanese en
C 142	21.6	51.4	122590	4	AC148346	AC148346 Medicago	c 215	21.2	50.5	1500	10	AEU44960	AEU44960 Japanese en
C 143	21.6	51.4	127054	12	AC163249	AC163249 Loxodonta	c 216	21.2	50.5	1500	10	AEU44964	AEU44964 Japanese en
C 144	21.6	51.4	127270	5	AC036144	AC036144 Homo sapi	c 217	21.2	50.5	1500	10	AEU44966	AEU44966 Japanese en
C 145	21.6	51.4	129081	12	AC163247	AC163247 Loxodonta	c 218	21.2	50.5	1500	10	AEU44967	AEU44967 Japanese en
C 146	21.6	51.4	144824	6	CR936848	CR936848 Mouse DNA	c 219	21.2	50.5	1500	10	AEU44968	AEU44968 Japanese en
C 147	21.6	51.4	149103	12	AC128053	AC128053 Rattus no	c 220	21.2	50.5	1500	10	AEU44969	AEU44969 Japanese en
C 148	21.6	51.4	168829	5	AC009941	AC009941 Homo sapi	c 221	21.2	50.5	1500	10	AEU44973	AEU44973 Japanese en
C 149	21.6	51.4	179591	5	AC093902	AC093902 Homo sapi	c 222	21.2	50.5	1500	10	AEU70395	AEU70395 Japanese en
C 150	21.6	51.4	183224	12	AC048363	AC048363 Homo sapi	c 223	21.2	50.5	1500	10	AEU70396	AEU70396 Japanese en
C 151	21.6	51.4	190469	6	AC109172	AC109172 Mus muscu	c 224	21.2	50.5	1500	10	AEU70399	AEU70399 Japanese en
C 152	21.6	51.4	207878	12	AC013566	AC013566 Homo sapi	c 225	21.2	50.5	1500	10	AEU70400	AEU70400 Japanese en
C 153	21.6	51.4	213518	12	AC165932	AC165932 Bos tauru	c 226	21.2	50.5	1500	10	AEU70405	AEU70405 Japanese en
C 154	21.6	51.4	221115	12	AC111437	AC111437 Rattus no	c 227	21.2	50.5	1500	10	AEU70411	AEU70411 Japanese en
C 155	21.6	51.4	226967	6	AL627425	AL627425 Mouse DNA	c 228	21.2	50.5	1500	10	AEU70412	AEU70412 Japanese en
C 156	21.6	51.4	233589	6	AL683824	AL683824 Mouse DNA	c 229	21.2	50.5	1500	10	AEU70421	AEU70421 Japanese en
C 157	21.6	51.4	237899	6	AC102622	AC102622 Mus muscu	c 230	21.2	50.5	1500	10	AEU70421	AEU70421 Japanese en
C 158	21.6	51.4	240349	12	AC096133	AC096133 Rattus no	c 231	21.2	50.5	1500	10	AEU70421	AEU70421 Japanese en
C 159	21.6	51.0	407	2	AX916898	AX916898 Sequence	c 232	21.2	50.5	1500	10	AEU70421	AEU70421 Japanese en
C 160	21.4	51.0	407	2	BD052431	BD052431 Sequence	c 233	21.2	50.5	2435	10	AEU21054	AEU21054 Japanese [
C 161	21.4	51.0	407	2	AR752692	AR752692 Sequence	c 234	21.2	50.5	2436	10	AEU03691	AEU03691 Japanese en
C 162	21.4	51.0	428	11	AY844413	AY844413 Hyia cycl	c 235	21.2	50.5	2436	10	AEU03692	AEU03692 Japanese en
C 163	21.4	51.0	428	11	AY844438	AY844438 Hyia neph	c 236	21.2	50.5	2436	10	AEU03694	AEU03694 Japanese en
C 164	21.4	51.0	428	11	AY844487	AY844487 Osteopiliu	c 237	21.2	50.5	2436	10	AEU21057	AEU21057 Japanese en



238	21.2	50.5	4406	10	AY027863	AY027863 Japanese	C 311	21	50.0	692	2	CQ411716	CQ411716 Sequence
239	21.2	50.5	5853	13	TRBANAT1	M20871 T. brucei va	C 312	21	50.0	758	7	BV563202	BV563202 qz274g09.
240	21.2	50.5	6234	4	AY622826	AY622826 Emericell	C 313	21	50.0	1008	5	AY453396	AY453396 Homo sapi
241	21.2	50.5	10951	10	JEVLINGCG	L78128 Japanese en	C 314	21	50.0	1008	5	BT007150	BT007150 Homo sapi
242	21.2	50.5	10970	10	AF254452	AF254452 Japanese	C 315	21	50.0	1008	5	BT020135	BT020135 Homo sapi
243	21.2	50.5	10970	10	AF254453	AF254453 Japanese	C 316	21	50.0	1008	8	AY888499	AY888499 Synthetic
244	21.2	50.5	10970	10	AY303791	AY303791 Japanese	C 317	21	50.0	1008	8	AY890300	AY890300 Synthetic
245	21.2	50.5	10970	10	AY303792	AY303792 Japanese	C 318	21	50.0	1008	8	AY892776	AY892776 Synthetic
246	21.2	50.5	10976	10	AF014160	AF014160 Japanese	C 319	21	50.0	1008	8	BT008192	BT008192 Synthetic
247	21.2	50.5	10976	10	AF014161	AF014161 Japanese	C 320	21	50.0	1096	2	CQ413562	CQ413562 Sequence
248	21.2	50.5	10976	10	AF098735	AF098735 Japanese	C 321	21	50.0	1386	2	CQ413739	CQ413739 Sequence
249	21.2	50.5	10976	10	AF098736	AF098736 Japanese	C 322	21	50.0	1580	5	AB028741	AB028741 Homo sapi
250	21.2	50.5	10976	10	AF098737	AF098737 Japanese	C 323	21	50.0	1586	5	BC014613	BC014613 Homo sapi
251	21.2	50.5	10976	10	AY508812	AY508812 Japanese	C 324	21	50.0	1587	14	BC102576	BC102576 Bos tauru
252	21.2	50.5	10976	10	AY508813	AY508813 Japanese	C 325	21	50.0	1599	6	BC034063	BC034063 Mus muscu
253	21.2	50.5	10977	10	AB196923	AB196923 Japanese	C 326	21	50.0	1648	2	AX405638	AX405638 Sequence
254	21.2	50.5	10977	10	AB196924	AB196924 Japanese	C 327	21	50.0	1936	6	BC036174	BC036174 Mus muscu
255	21.2	50.5	10977	10	AB196925	AB196925 Japanese	C 328	21	50.0	2102	6	BC039217	BC039217 Mus muscu
256	21.2	50.5	10977	10	AB196926	AB196926 Japanese	C 329	21	50.0	2145	2	CQ413719	CQ413719 Sequence
257	21.2	50.5	10977	10	AF069076	AF069076 Japanese	C 330	21	50.0	2174	6	BC036173	BC036173 Mus muscu
258	21.2	50.5	10977	10	AF080251	AF080251 Japanese	C 331	21	50.0	2420	6	AY079517S2	AY079517 Mus muscu
259	21.2	50.5	10977	10	AF486638	AF486638 Japanese	C 332	21	50.0	12027	15	AE008569	AE008569 Rickettsi
260	21.2	50.5	30221	12	AC014454	AC014454 Drosophil	C 333	21	50.0	96589	2	AX695599	AX695599 Sequence
261	21.2	50.5	30508	13	TRBESAG	L20156 Trypanosoma	C 334	21	50.0	103667	12	AC178833	AC178833 Strongylio
262	21.2	50.5	40592	5	HSL191F1	Z68756 Human DNA s	C 335	21	50.0	110000	15	CP000076_14	Continuation (15 o
263	21.2	50.5	56016	12	AC091679	AC091679 Mus muscu	C 336	21	50.0	125517	5	HS191156	AL009050 Human DNA
264	21.2	50.5	84265	11	AB028136	AB028136 Gallus ga	C 337	21	50.0	126080	5	AY029472	AY029472 Homo sapi
265	21.2	50.5	93871	12	AC165216	AC165216 Medicago	C 338	21	50.0	137723	5	AC110814	AC110814 Homo sapi
266	21.2	50.5	123706	13	TH25N7	AL671259 Trypanoso	C 339	21	50.0	151019	12	AC151938	AC151938 Dasyypu n
267	21.2	50.5	137759	13	TBN19B2	AL671256 Trypanoso	C 340	21	50.0	152680	12	AC173490	AC173490 Bos tauru
268	21.2	50.5	148365	12	AC165219	AC165219 Medicago	C 341	21	50.0	158793	12	AC169239	AC169239 Bos tauru
269	21.2	50.5	157119	5	AL3390065	AL3390065 Human DNA	C 342	21	50.0	159791	5	AC009951	AC009951 Homo sapi
270	21.2	50.5	158822	12	AC136174	AC136174 Rattus no	C 343	21	50.0	164281	11	CR318673	CR318673 Zebrafish
271	21.2	50.5	161807	12	BX936424	BX936424 Danio rer	C 344	21	50.0	166469	12	CT009633	CT009633 Danio rer
272	21.2	50.5	163323	6	AC091395	AC091395 Mus muscu	C 345	21	50.0	169516	12	AC013934	AC013934 Drosophil
273	21.2	50.5	163690	12	AC140908	AC140908 Homo sapi	C 346	21	50.0	170662	13	AC009905	AC009905 Drosophil
274	21.2	50.5	164238	5	AC106745	AC106745 Homo sapi	C 347	21	50.0	175795	12	AC155028	AC155028 Bos tauru
275	21.2	50.5	164271	6	AC122515	AC122515 Mus muscu	C 348	21	50.0	182128	11	AC147446	AC147446 Gallus ga
276	21.2	50.5	164386	13	AC006491	AC006491 Drosophil	C 349	21	50.0	184341	12	AC135835	AC135835 Papio anu
277	21.2	50.5	170298	13	AE003690	AE003690 Drosophil	C 350	21	50.0	190196	12	AC129837	AC129837 Papio anu
278	21.2	50.5	173561	12	AC016809	AC016809 Homo sapi	C 351	21	50.0	194763	12	AC169116	AC169116 Bos tauru
279	21.2	50.5	174471	6	AC122534	AC122534 Mus muscu	C 352	21	50.0	195466	12	AC176508	AC176508 Strongylio
280	21.2	50.5	192226	5	AC109000	AC109000 Rattus no	C 353	21	50.0	199299	6	AC131718	AC131718 Mus muscu
281	21.2	50.5	192758	5	AC104434	AC104434 Homo sapi	C 354	21	50.0	202399	5	AC090966	AC090966 Papio anu
282	21.2	50.5	202844	6	AC110211	AC110211 Mus muscu	C 355	21	50.0	206846	12	AC165513	AC165513 Bos tauru
283	21.2	50.5	203133	12	AC138168	AC138168 Sus scrof	C 356	21	50.0	213473	6	AC140287	AC140287 Mus muscu
284	21.2	50.5	204001	12	AC148896	AC148896 Otlemur	C 357	21	50.0	216248	12	AC168788	AC168788 Strongylio
285	21.2	50.5	206098	12	AC156915	AC156915 Bos tauru	C 358	21	50.0	217742	12	AC172111	AC172111 Macaca mu
286	21.2	50.5	207639	6	AL844591	AL844591 Mouse DNA	C 359	21	50.0	226088	12	AC152967	AC152967 Bos tauru
287	21.2	50.5	208709	12	AC154058	AC154058 Otlemur	C 360	21	50.0	234563	12	AC123093	AC123093 Rattus no
288	21.2	50.5	217359	12	AC151059	AC151059 Bos tauru	C 361	21	50.0	244141	12	AC096072	AC096072 Rattus no
289	21.2	50.5	217991	12	AC130504	AC130504 Rattus no	C 362	21	50.0	258290	12	AC109715	AC109715 Rattus no
290	21.2	50.5	220691	6	AL672270	AL672270 Mouse DNA	C 363	21	50.0	258558	13	AE003654	AE003654 Drosophil
291	21.2	50.5	226284	6	AC099609	AC099609 Mus muscu	C 364	21	50.0	264683	12	AC105646	AC105646 Rattus no
292	21.2	50.5	231604	12	AC130095	AC130095 Rattus no	C 365	21	50.0	267565	12	AC098917	AC098917 Rattus no
293	21.2	50.5	244899	12	AC110931	AC110931 Rattus no	C 366	21	50.0	300953	12	AC162349	AC162349 Bos tauru
294	21.2	50.5	245387	12	AL358052	AL358052 Homo sapi	C 367	21	50.0	334392	12	AC118975	AC118975 Rattus no
295	21.2	50.5	247418	12	AC107347	AC107347 Rattus no	C 368	21	50.0	348764	15	BX569689	BX569689 Synchoco
296	21.2	50.5	253919	12	AC109035	AC109035 Rattus no	C 369	20.8	49.5	917	14	DQ240470	DQ240470 Nycereut
297	21.2	50.5	286014	12	AC112125	AC112125 Rattus no	C 370	20.8	49.5	1728	13	MSU45682	AR348475 Sequence
298	21.2	50.5	302156	12	AC157159	AC157159 Bos tauru	C 371	20.8	49.5	1832	13	AR348475	U46682 Manduca sex
299	21.2	50.5	302156	12	AC110456	AC110456 Rattus no	C 372	20.8	49.5	1935	5	AK098749	AK098749 Homo sapi
300	21	50.0	251	15	AY345080	AY345080 Rickettsi	C 373	20.8	49.5	4001	5	AC133795	AC133795 Homo sapi
301	21	50.0	428	11	AY844463	AY844463 Hyla taen	C 374	20.8	49.5	5565	5	AF166009S3	AF166011 Homo sapi
302	21	50.0	527	2	CQ396550	CQ396550 Sequence	C 375	20.8	49.5	89295	12	AC159736	AC159736 Homo sapi
303	21	50.0	527	2	CQ402864	CQ402864 Sequence	C 376	20.8	49.5	101034	5	AC096550	AC096550 Bos tauru
304	21	50.0	558	7	BV586363	BV586363 G591P6319	C 377	20.8	49.5	110000	12	PFMAL8P1_05	Continuation (6 of
305	21	50.0	559	2	CQ398698	CQ398698 Sequence	C 378	20.8	49.5	110000	15	AP007255_30	Continuation (31 o
306	21	50.0	559	2	CQ404981	CQ404981 Sequence	C 379	20.8	49.5	110000	15	AP007255_31	Continuation (32 o
307	21	50.0	579	2	CQ397122	CQ397122 Sequence	C 380	20.8	49.5	110828	12	AC087586	AC087586 Homo sapi
308	21	50.0	579	2	CQ403428	CQ403428 Sequence	C 381	20.8	49.5	114950	5	AC097652	AC097652 Homo sapi
309	21	50.0	620	7	BV048212	BV048212 S212P6044	C 382	20.8	49.5	120769	14	CR956354	CR956354 Wallaby D
310	21	50.0	625	2	CQ411365	CQ411365 Sequence	C 383	20.8	49.5	123716	12	CT027747	Danio rer



C 530	20.4	48.6	264	2	AR406937	Sequence	AR406937	20.4	48.6	162861	12	AC018597
C 531	20.4	48.6	264	2	AR440787	Sequence	AR440787	20.4	48.6	164168	5	AL135927
C 532	20.4	48.6	264	2	AR472945	Sequence	AR472945	20.4	48.6	164179	5	AC007227
C 533	20.4	48.6	264	2	AR543598	Sequence	AR543598	20.4	48.6	164450	13	AC011763
C 534	20.4	48.6	264	2	AR638796	Sequence	AR638796	20.4	48.6	164955	12	AC131864
C 535	20.4	48.6	264	2	AR695429	Sequence	AR695429	20.4	48.6	167647	6	AC144913
C 536	20.4	48.6	264	2	AX063196	Sequence	AX063196	20.4	48.6	169085	6	AC132255
C 537	20.4	48.6	264	2	AX368113	Sequence	AX368113	20.4	48.6	169085	6	AC102666
C 538	20.4	48.6	363	2	AR622535	Sequence	AR622535	20.4	48.6	171875	12	AC099746
C 539	20.4	48.6	540	7	AU027881	Rattus no	AU027881	20.4	48.6	172382	12	AC149046
C 540	20.4	48.6	615	7	BV352364	S230P6361	BV352364	20.4	48.6	172659	12	AC024320
C 541	20.4	48.6	629	2	CQ523848	Sequence	CQ523848	20.4	48.6	172755	5	CNS01RGL
C 542	20.4	48.6	703	7	BV052360	S212P6351	BV052360	20.4	48.6	173403	12	AC164861
C 543	20.4	48.6	707	2	AX436879	Sequence	AX436879	20.4	48.6	173891	12	BX927251
C 544	20.4	48.6	708	2	AX436719	Sequence	AX436719	20.4	48.6	176105	12	AC153235
C 545	20.4	48.6	728	2	BD148430	Primer fo	BD148430	20.4	48.6	176399	5	CNS059E1
C 546	20.4	48.6	728	2	AX868368	Sequence	AX868368	20.4	48.6	177041	5	AC126024
C 547	20.4	48.6	770	11	CR405907	Gallus ga	CR405907	20.4	48.6	177273	5	AC074194
C 548	20.4	48.6	1356	2	CQ717415	Sequence	CQ717415	20.4	48.6	179576	12	AC170664
C 549	20.4	48.6	1752	14	BC110126	Bos tauru	BC110126	20.4	48.6	179738	5	AL1354938
C 550	20.4	48.6	2523	2	AX714593	Sequence	AX714593	20.4	48.6	181120	6	AC161148
C 551	20.4	48.6	2523	5	AK057043	Homo sapi	AK057043	20.4	48.6	181782	6	AC161172
C 552	20.4	48.6	2648	2	AR717849	Sequence	AR717849	20.4	48.6	188137	12	AC148796
C 553	20.4	48.6	2648	2	AX746824	Sequence	AX746824	20.4	48.6	193044	12	AC027769
C 554	20.4	48.6	2648	5	AK091168	Homo sapi	AK091168	20.4	48.6	193312	12	AC157839
C 555	20.4	48.6	2655	5	BC103826	Homo sapi	BC103826	20.4	48.6	194339	6	AC109199
C 556	20.4	48.6	2655	5	BC103827	Homo sapi	BC103827	20.4	48.6	196972	12	AC166531
C 557	20.4	48.6	2737	2	CQ606770	Sequence	CQ606770	20.4	48.6	197273	6	AC024607
C 558	20.4	48.6	2952	5	BC008928	Homo sapi	BC008928	20.4	48.6	198777	11	AL807749
C 559	20.4	48.6	2970	5	BC064624	Homo sapi	BC064624	20.4	48.6	205550	12	AL148945
C 560	20.4	48.6	2973	5	BC075854	Homo sapi	BC075854	20.4	48.6	208470	6	AL928883
C 561	20.4	48.6	2998	5	BSA409349	Homo sapi	BSA409349	20.4	48.6	214809	12	AC099407
C 562	20.4	48.6	3057	2	CS245209	Sequence	CS245209	20.4	48.6	216158	12	AC113563
C 563	20.4	48.6	3240	2	BD156445	Primer fo	BD156445	20.4	48.6	216289	6	AC123712
C 564	20.4	48.6	3240	2	AX877025	Sequence	AX877025	20.4	48.6	216615	6	AC107850
C 565	20.4	48.6	3240	5	AK001574	Homo sapi	AK001574	20.4	48.6	217725	13	AE003717
C 566	20.4	48.6	3722	5	AK124755	Homo sapi	AK124755	20.4	48.6	218524	6	AC151721
C 567	20.4	48.6	5336	11	AB097142	Danio rer	AB097142	20.4	48.6	224009	12	AC164301
C 568	20.4	48.6	6172	15	MTMCR	M. thermoau	MTMCR	20.4	48.6	224101	6	AC145301
C 569	20.4	48.6	21295	2	AR619799	Sequence	AR619799	20.4	48.6	227407	12	AC170467
C 570	20.4	48.6	56179	4	AP004626	Lotus cor	AP004626	20.4	48.6	228155	6	AC164411
C 571	20.4	48.6	67395	12	AC134417	Mus muscu	AC134417	20.4	48.6	228761	12	AC123314
C 572	20.4	48.6	67971	5	AL138124	Homo sapi	AL138124	20.4	48.6	230187	12	AC152198
C 573	20.4	48.6	85763	5	AL359455	Human DNA	AL359455	20.4	48.6	230212	6	CT025668
C 574	20.4	48.6	92969	6	BX842667	Mouse DNA	BX842667	20.4	48.6	231337	12	CR293502
C 575	20.4	48.6	94135	12	AC168633	Strongylo	AC168633	20.4	48.6	241100	12	AC095453
C 576	20.4	48.6	110000	2	BD430793_16	Continuation (17 o	BD430793_16	20.4	48.6	241958	12	AC162594
C 577	20.4	48.6	110000	4	AP008213_236	Continuation (237	AP008213_236	20.4	48.6	247017	12	AC094921
C 578	20.4	48.6	110000	4	AP008214_090	Continuation (91 o	AP008214_090	20.4	48.6	251796	12	AC094921
C 579	20.4	48.6	110000	4	CR382130_25	Continuation (26 o	CR382130_25	20.4	48.6	251806	12	AC096305
C 580	20.4	48.6	110000	4	CR382131_18	Continuation (19 o	CR382131_18	20.4	48.6	252995	12	AC091638
C 581	20.4	48.6	110000	12	AC013623_1	Continuation (2 of	AC013623_1	20.4	48.6	255019	12	AC162413
C 582	20.4	48.6	110000	12	AC021632_0	Continuation (30 o	AC021632_0	20.4	48.6	263704	12	AC162592
C 583	20.4	48.6	110000	15	CR954246_29	Continuation (17 o	CR954246_29	20.4	48.6	263776	12	AC087335
C 584	20.4	48.6	119235	5	AL011817	Homo sapi	AL011817	20.4	48.6	264661	12	AC094217
C 585	20.4	48.6	122864	5	AL390863	Human DNA	AL390863	20.4	48.6	265118	2	AX041922
C 586	20.4	48.6	123787	4	AP003747	Oryza sat	AP003747	20.4	48.6	265118	15	CNSPAX06
C 587	20.4	48.6	128318	5	AP007450	Homo sapi	AP007450	20.4	48.6	276285	12	AC128520
C 588	20.4	48.6	129108	12	AC162742	Loxodonta	AC162742	20.4	48.6	280325	12	AC112293
C 589	20.4	48.6	133248	4	AC163974	Loxodonta	AC163974	20.4	48.6	319246	12	BX546472
C 590	20.4	48.6	141473	11	BX942819	Zebraphis	BX942819	20.4	48.1	384	2	AX537754
C 591	20.4	48.6	145988	12	AC073219	Homo sapi	AC073219	20.2	48.1	384	4	MR4400724
C 592	20.4	48.6	153106	12	AC162469	Bos tauru	AC162469	20.2	48.1	428	11	AY844365
C 593	20.4	48.6	155544	4	AP005799	Oryza sat	AP005799	20.2	48.1	428	11	AY844366
C 594	20.4	48.6	156318	5	AC132834	Homo sapi	AC132834	20.2	48.1	428	11	AY844377
C 595	20.4	48.6	157020	12	AC074118	Homo sapi	AC074118	20.2	48.1	428	11	AY844384
C 596	20.4	48.6	157660	12	AC027774	Homo sapi	AC027774	20.2	48.1	428	11	AY844385
C 597	20.4	48.6	157905	4	AP005495	Oryza sat	AP005495	20.2	48.1	428	11	AY844386
C 598	20.4	48.6	159144	12	AC116649	Homo sapi	AC116649	20.2	48.1	428	11	AY844388
C 599	20.4	48.6	160355	12	AC022644	Homo sapi	AC022644	20.2	48.1	428	11	AY844392
C 600	20.4	48.6	160561	12	AC017724	Drosophill	AC017724	20.2	48.1	428	11	AY844394
C 601	20.4	48.6	160990	5	AL157702	Human DNA	AL157702	20.2	48.1	428	11	AY844396
C 602	20.4	48.6	162028	12	AC149108	Papio anu	AC149108	20.2	48.1	428	11	AY844400

676	20.2	48.1	428	11	AY8444402	AY8444402	Hyia call	749	20.2	48.1	110000	15	AE000657_00	AE000657 Aquifex a
677	20.2	48.1	428	11	AY8444404	AY8444404	Hyia carn	c 750	20.2	48.1	110000	15	CP000103_06	Continuation (7 of
678	20.2	48.1	428	11	AY8444408	AY8444408	Hyia cine	751	20.2	48.1	110000	15	CP000112_06	Continuation (7 of
679	20.2	48.1	428	11	AY8444409	AY8444409	Hyia circ	752	20.2	48.1	110000	15	AE008384_04	Continuation (5 of
680	20.2	48.1	428	11	AY8444410	AY8444410	Hyia coly	c 753	20.2	48.1	111868	12	AC180846	AC180846 Strongylo
681	20.2	48.1	428	11	AY8444417	AY8444417	Hyia gies	754	20.2	48.1	111868	12	AC092078	AC092078 Homo sapi
682	20.2	48.1	428	11	AY8444418	AY8444418	Hyia grat	c 755	20.2	48.1	116194	12	AC172875	AC172875 Brassica
683	20.2	48.1	428	11	AY8444419	AY8444419	Hyia hyla	756	20.2	48.1	119420	12	AC027485	AC027485 Homo sapi
684	20.2	48.1	428	11	AY8444423	AY8444423	Hyia lema	c 757	20.2	48.1	123009	12	AC181492	AC181492 Strongylo
685	20.2	48.1	428	11	AY8444425	AY8444425	Hyia leuc	c 758	20.2	48.1	123291	5	AC104070	AC104070 Homo sapi
686	20.2	48.1	428	11	AY8444430	AY8444430	Hyia micr	759	20.2	48.1	124552	5	AC011403	AC011403 Homo sapi
687	20.2	48.1	428	11	AY8444435	AY8444435	Hyia miya	c 760	20.2	48.1	131152	6	AL929068	AL929068 Mouse DNA
688	20.2	48.1	428	11	AY8444436	AY8444436	Hyia mult	c 761	20.2	48.1	134420	5	AC010428	AC010428 Homo sapi
689	20.2	48.1	428	11	AY8444439	AY8444439	Hyia palm	c 762	20.2	48.1	134940	12	AC018939	AC018939 Homo sapi
690	20.2	48.1	428	11	AY8444440	AY8444440	Hyia parv	763	20.2	48.1	135028	12	AC181466	AC181466 Strongylo
691	20.2	48.1	428	11	AY8444444	AY8444444	Hyia pseu	764	20.2	48.1	136853	6	AC165307	AC165307 Mus muscu
692	20.2	48.1	428	11	AY8444446	AY8444446	Hyia rhod	c 765	20.2	48.1	138159	5	AL138965	AL138965 Human DNA
693	20.2	48.1	428	11	AY8444451	AY8444451	Hyia sara	c 766	20.2	48.1	140042	6	AC157909	AC157909 Mus muscu
694	20.2	48.1	428	11	AY8444455	AY8444455	Hyia sib1	c 767	20.2	48.1	143092	13	CEY1797B	AL023828 Caenorhab
695	20.2	48.1	428	11	AY8444456	AY8444456	Hyia sp.	c 768	20.2	48.1	143461	12	AC141021	AC141021 Rattus no
696	20.2	48.1	428	11	AY8444457	AY8444457	Hyia aff.	769	20.2	48.1	146699	5	BS000226	BS000226 Pan trogl
697	20.2	48.1	428	11	AY8444458	AY8444458	Hyia sp.	c 770	20.2	48.1	146891	12	AL845302	AL845302 Oryctolag
698	20.2	48.1	428	11	AY8444461	AY8444461	Hyia aff.	c 771	20.2	48.1	147219	11	AL845302	AL845302 Zebrafish
699	20.2	48.1	428	11	AY8444467	AY8444467	Hyia weyg	772	20.2	48.1	148628	12	AC157451	AC157451 Dasypus n
700	20.2	48.1	428	11	AY8444478	AY8444478	Osteoceph	773	20.2	48.1	148892	4	AP003235	AP003235 Oryza sat
701	20.2	48.1	428	11	AY8444481	AY8444481	Osteoceph	774	20.2	48.1	152523	5	AC007009	AC007009 Homo sapi
702	20.2	48.1	428	11	AY8444483	AY8444483	Osteoceph	c 775	20.2	48.1	154524	6	AC132278	AC132278 Mus muscu
703	20.2	48.1	428	11	AY8444484	AY8444484	Osteoceph	776	20.2	48.1	156381	12	AC171966	AC171966 Bos tauru
704	20.2	48.1	428	11	AY8444485	AY8444485	Osteoceph	c 777	20.2	48.1	157295	12	AC079055	AC079055 Homo sapi
705	20.2	48.1	428	11	AY8444490	AY8444490	Phrynohya	c 778	20.2	48.1	159469	6	AC121920	AC121920 Mus muscu
706	20.2	48.1	428	11	AY8444493	AY8444493	Phrynohya	c 779	20.2	48.1	162020	12	AC098864	AC098864 Homo sapi
707	20.2	48.1	428	11	AY8444518	AY8444518	Scinax el	c 780	20.2	48.1	163755	6	AC168570	AC168570 Mus muscu
708	20.2	48.1	428	11	AY8444520	AY8444520	Scinax na	c 781	20.2	48.1	164455	6	AC102466	AC102466 Mus muscu
709	20.2	48.1	428	11	AY8444521	AY8444521	Scinax ru	782	20.2	48.1	164686	5	CNS057CG	AL350597 Human chr
710	20.2	48.1	428	11	AY8444522	AY8444522	Scinax sq	783	20.2	48.1	165030	12	AC179774	AC179774 Strongylo
711	20.2	48.1	428	11	AY8444523	AY8444523	Scinax st	c 784	20.2	48.1	165030	12	AC179774	AC179774 Strongylo
712	20.2	48.1	428	11	AY8444530	AY8444530	Tepuihya	c 785	20.2	48.1	165175	12	AC168383	AC168383 Strongylo
713	20.2	48.1	428	11	AY8444531	AY8444531	Trachycep	c 786	20.2	48.1	165880	5	AL136372	AL136372 Human DNA
714	20.2	48.1	922	5	AF0699367	AF0699367	Hylobates	c 787	20.2	48.1	166407	12	AC176375	AC176375 Strongylo
715	20.2	48.1	1018	5	AY0121334	AY0121334	Hylobates	788	20.2	48.1	166451	5	AC018994	AC018994 Homo sapi
716	20.2	48.1	1455	4	BT0002026	BT0002026	Arabis	789	20.2	48.1	167144	6	AC158748	AC158748 Mus muscu
717	20.2	48.1	1534	11	AY323766	AY323766	Hyia cine	c 790	20.2	48.1	167664	12	AC073870	AC073870 Homo sapi
718	20.2	48.1	1545	2	AX300259	AX300259	Sequence	791	20.2	48.1	170393	12	AC024304	AC024304 Homo sapi
719	20.2	48.1	1557	4	AY057647	AY057647	Arabis	c 792	20.2	48.1	170936	5	AC147028	AC147028 Pan trogl
720	20.2	48.1	1733	4	AY136392	AY136392	Arabis	793	20.2	48.1	171790	12	AC011106	AC011106 Homo sapi
721	20.2	48.1	4327	6	BC079620	BC079620	Mus muscu	c 794	20.2	48.1	172576	5	CNS01DSG	AL121775 Human chr
722	20.2	48.1	5919	6	BC095958	BC095958	Mus muscu	795	20.2	48.1	172949	12	AC123431	AC123431 Rattus no
723	20.2	48.1	6005	6	AY043226	AY043226	Rattus no	c 796	20.2	48.1	173094	12	AC141547	AC141547 Rattus no
724	20.2	48.1	6195	6	BC058681	BC058681	Mus muscu	c 797	20.2	48.1	174130	11	AL732628	AL732628 Zebrafish
725	20.2	48.1	6268	6	AK122284	AK122284	Mus muscu	798	20.2	48.1	176190	12	AC180851	AC180851 Strongylo
726	20.2	48.1	6400	6	AF026504	AF026504	Rattus no	799	20.2	48.1	176269	12	AC012663	AC012663 Homo sapi
727	20.2	48.1	6429	6	AK129343	AK129343	Mus muscu	c 800	20.2	48.1	176933	5	AC011994	AC011994 Homo sapi
728	20.2	48.1	16472	5	HLMITCSEQ	X99256	Hylobates l	801	20.2	48.1	177255	12	AC123235	AC123235 Rattus no
729	20.2	48.1	4096	13	U80437	U80437	Caenorhabdi	c 802	20.2	48.1	178830	4	AP003286	AP003286 Oryza sat
730	20.2	48.1	49799	12	AC100909	AC100909	Mus muscu	c 803	20.2	48.1	179491	12	AC020673	AC020673 Homo sapi
731	20.2	48.1	53394	5	AC013276	AC013276	Homo sapi	c 804	20.2	48.1	180363	6	AC154755	AC154755 Mus muscu
732	20.2	48.1	68316	12	CEH16D04	Z92795	Caenorhabdi	c 805	20.2	48.1	181208	6	AC124370	AC124370 Mus muscu
733	20.2	48.1	71260	11	AC151318	AC151318	Xenopus t	c 806	20.2	48.1	182108	12	AC170644	AC170644 Bos tauru
734	20.2	48.1	71612	12	AC173273	AC173273	Bos tauru	c 807	20.2	48.1	183494	12	AC079989	AC079989 Rattus no
735	20.2	48.1	71739	5	AL391360	AL391360	Human DNA	c 808	20.2	48.1	184327	6	AC120859	AC120859 Mus muscu
736	20.2	48.1	77101	5	AC090684	AC090684	Homo sapi	c 809	20.2	48.1	184377	6	AC121579	AC121579 Mus muscu
737	20.2	48.1	79186	12	AC006781	AC006781	Caenorhab	810	20.2	48.1	184439	12	AP003085	AP003085 Homo sapi
738	20.2	48.1	79186	12	AC006859	AC006859	Caenorhab	811	20.2	48.1	184623	6	AC158570	AC158570 Mus muscu
739	20.2	48.1	79452	12	AC016536	AC016536	Homo sapi	812	20.2	48.1	185561	12	AC117124	AC117124 Rattus no
740	20.2	48.1	95449	5	AP001620	AP001620	Homo sapi	c 813	20.2	48.1	185809	12	AC141579	AC141579 Rattus no
741	20.2	48.1	97990	5	AC025167	AC025167	Homo sapi	c 814	20.2	48.1	187451	12	AC176132	AC176132 Strongylo
742	20.2	48.1	103244	14	CR956378	CR956378	Pig DNA s	c 815	20.2	48.1	187568	5	AP002967	AP002967 Homo sapi
743	20.2	48.1	103917	12	AC125501	AC125501	Takifugu	c 816	20.2	48.1	187752	6	AC124689	AC124689 Mus muscu
744	20.2	48.1	105104	12	AC110828	Continuation (7 of	c 817	20.2	48.1	188376	6	AC112267	AC112267 Mus muscu	
745	20.2	48.1	105259	5	AC116331	Continuation (7 of	c 818	20.2	48.1	189137	6	AC079378	AC079378 Rattus no	
746	20.2	48.1	105543	4	ATAC09325	AC09325	Arabis	c 819	20.2	48.1	189847	12	AC091964	AC091964 Homo sapi
747	20.2	48.1	108556	12	AC170551	AC170551	Bos tauru	c 820	20.2	48.1	190374	12	AC132647	AC132647 Rattus no
748	20.2	48.1	110000	4	AP008207_376	Continuation (377	c 821	20.2	48.1	191098	12	AC178716	AC178716 Strongylo	

822	20.2	48.1	191747	12	AC113675	AC113675 Rattus no	895	20	47.6	542	2	AX105689	AX105689 Sequence
823	20.2	48.1	193412	12	AC170376	AC170376 Bos tauru	896	20	47.6	546	2	AX319688	AX319688 Sequence
824	20.2	48.1	195269	6	AC153359	AC153359 Mus muscu	897	20	47.6	576	2	AR035907	AR035907 Sequence
825	20.2	48.1	196898	12	AC139945	AC139945 Rattus no	898	20	47.6	576	2	I20143	I20143 Sequence 49
826	20.2	48.1	198328	12	AC167488	AC167488 Bos tauru	899	20	47.6	576	2	AR340321	AR340321 Sequence
827	20.2	48.1	198392	6	AC022236	AC022236 Mus muscu	900	20	47.6	576	10	HFCPE1R	L16645 Hepatitis C
828	20.2	48.1	199746	12	AC173610	AC173610 Bos tauru	901	20	47.6	624	14	SSSP8C23	QJ000020 Sus scrof
829	20.2	48.1	199853	12	AC073785	AC073785 Mus muscu	902	20	47.6	649	2	CQ409812	CQ409812 Sequence
830	20.2	48.1	199860	6	AC102131	AC102131 Mus muscu	903	20	47.6	808	7	BV070218	BV070218 S21296040
831	20.2	48.1	200353	6	AC166575	AC166575 Mus muscu	904	20	47.6	1385	4	AK120666	AK120666 Oryza sat
832	20.2	48.1	201694	6	AC111115	AC111115 Mus muscu	905	20	47.6	4636	4	AK120507	AK120507 Oryza sat
833	20.2	48.1	201721	12	AC157290	AC157290 Bos tauru	906	20	47.6	7690	2	AX346025	AX346025 Sequence
834	20.2	48.1	203187	12	CT485788	CT485788 Mus muscu	907	20	47.6	9388	5	AB095937	AB095937 Homo sapi
835	20.2	48.1	205940	12	AC021830	AC021830 Homo sapi	908	20	47.6	24387	5	HSJ185D5	AL118498 Human DNA
836	20.2	48.1	206075	12	AC118511	AC118511 Rattus no	909	20	47.6	38345	5	AY055383	AY055383 Homo sapi
837	20.2	48.1	206383	6	AC008160	AC008160 Mus muscu	910	20	47.6	38703	5	AC000031	AC000031 Homo sapi
838	20.2	48.1	207922	6	AL450399	AL450399 Mouse DNA	911	20	47.6	42659	12	AC165608	AC165608 Bos tauru
839	20.2	48.1	208332	6	AL450399	AL450399 Mouse DNA	912	20	47.6	57358	15	AY803022	AY803022 Neisseria
840	20.2	48.1	209432	12	AC111308	AC111308 Rattus no	913	20	47.6	77227	12	AC025099	AC025099 Homo sapi
841	20.2	48.1	209885	12	AC172289	AC172289 Bos tauru	914	20	47.6	78794	4	OSJN00123	AL606996 Oryza sat
842	20.2	48.1	209999	12	AC019128	AC019128 Homo sapi	915	20	47.6	82767	11	AL672065	AL672065 Zebrafish
843	20.2	48.1	212104	5	AC012497	AC012497 Homo sapi	916	20	47.6	97572	4	AP004155	AP004155 Oryza sat
844	20.2	48.1	213047	12	AC020557	AC020557 Homo sapi	917	20	47.6	99766	12	AL353702	AL353702 Homo sapi
845	20.2	48.1	213181	12	AC122949	AC122949 Rattus no	918	20	47.6	100999	10	AF270937	AF270937 Plutella
846	20.2	48.1	213656	6	AC166172	AC166172 Mus muscu	919	20	47.6	102606	12	AC176436	AC176436 Strongylo
847	20.2	48.1	213669	6	AC144940	AC144940 Mus muscu	920	20	47.6	104353	5	HS1170K4	AL022314 Human DNA
848	20.2	48.1	214372	12	AC148959	AC148959 Otolomur	921	20	47.6	108514	12	AC084722	AC084722 Homo sapi
849	20.2	48.1	216380	6	AC113113	AC113113 Mus muscu	922	20	47.6	109540	4	AC147364	AC147364 Medicago
850	20.2	48.1	219599	12	AC109416	AC109416 Rattus no	923	20	47.6	110000	4	AP008214	Continuation 195 o
851	20.2	48.1	222545	12	AC109051	AC109051 Rattus no	924	20	47.6	110000	4	AP008214	Continuation 185 o
852	20.2	48.1	226048	6	AC074042	AC074042 Mus muscu	925	20	47.6	110000	15	BA000040	Continuation (20 o
853	20.2	48.1	226153	12	AC073704	AC073704 Mus muscu	926	20	47.6	117580	12	AP006455	AP006455 Oryza sat
854	20.2	48.1	227202	12	AC128501	AC128501 Rattus no	927	20	47.6	121581	12	AC167325	AC167325 Loxodonta
855	20.2	48.1	227492	6	AC116585	AC116585 Mus muscu	928	20	47.6	123375	4	OSJN00195	AL662933 Oryza sat
856	20.2	48.1	231435	12	AC113813	AC113813 Rattus no	929	20	47.6	129489	6	AL713860	AL713860 Mouse DNA
857	20.2	48.1	231485	12	AC098427	AC098427 Rattus no	930	20	47.6	136224	6	AC139062	AC139062 Mus muscu
858	20.2	48.1	232159	12	AC182457	AC182457 Gallus ga	931	20	47.6	137472	12	AC180955	AC180955 Strongylo
859	20.2	48.1	232483	12	AC129369	AC129369 Rattus no	932	20	47.6	140765	5	AC021301	AC021301 Homo sapi
860	20.2	48.1	233444	12	AC097829	AC097829 Rattus no	933	20	47.6	141628	5	AL391724	AL391724 Human DNA
861	20.2	48.1	234418	12	AC166379	AC166379 Bos tauru	934	20	47.6	143148	6	AL607034	AL607034 Mouse DNA
862	20.2	48.1	234584	12	AC111604	AC111604 Rattus no	935	20	47.6	146976	6	AC171199	AC171199 Mus muscu
863	20.2	48.1	236698	12	AC098997	AC098997 Rattus no	936	20	47.6	151468	6	AC154184	AC154184 Mus muscu
864	20.2	48.1	238953	12	AC166966	AC166966 Mus muscu	937	20	47.6	153938	12	AC168572	AC168572 Strongylo
865	20.2	48.1	239105	12	AC096791	AC096791 Rattus no	938	20	47.6	155595	12	AC078964	AC078964 Homo sapi
866	20.2	48.1	239591	12	AC128432	AC128432 Rattus no	939	20	47.6	156175	5	AC079807	AC079807 Homo sapi
867	20.2	48.1	240250	12	AC154886	AC154886 Bos tauru	940	20	47.6	159311	12	AC141040	AC141040 Rattus no
868	20.2	48.1	243366	12	AC126872	AC126872 Rattus no	941	20	47.6	160023	12	AC154256	AC154256 Mus muscu
869	20.2	48.1	243739	12	AC109679	AC109679 Rattus no	942	20	47.6	162362	6	AC154685	AC154685 Mus muscu
870	20.2	48.1	243770	12	AC123120	AC123120 Rattus no	943	20	47.6	165090	5	AC010655	AC010655 Homo sapi
871	20.2	48.1	245053	12	AC165130	AC165130 Bos tauru	944	20	47.6	166622	11	BX005135	BX005135 Zebrafish
872	20.2	48.1	245339	12	AC121656	AC121656 Rattus no	945	20	47.6	166669	5	AC113191	AC113191 Homo sapi
873	20.2	48.1	248580	12	AC117158	AC117158 Rattus no	946	20	47.6	167691	5	HS406A7	AL023284 Human DNA
874	20.2	48.1	249840	12	AC164639	AC164639 Mus muscu	947	20	47.6	167749	12	AC021489	AC021489 Homo sapi
875	20.2	48.1	251882	12	AC098075	AC098075 Rattus no	948	20	47.6	168863	5	AC011286	AC011286 Homo sapi
876	20.2	48.1	252054	12	AC097119	AC097119 Rattus no	949	20	47.6	170090	12	AC113593	AC113593 Mus muscu
877	20.2	48.1	252825	12	AC132024	AC132024 Rattus no	950	20	47.6	170928	6	AC119121	AC119121 Rattus no
878	20.2	48.1	258276	12	AC130846	AC130846 Rattus no	951	20	47.6	171162	12	AC068684	AC068684 Homo sapi
879	20.2	48.1	280025	12	AC108655	AC108655 Rattus no	952	20	47.6	171663	6	AC151475	AC151475 Mus muscu
880	20.2	48.1	280215	12	AC129648	AC129648 Rattus no	953	20	47.6	172449	5	AC146031	AC146031 Pan trogl
881	20.2	48.1	283806	12	AC167875	AC167875 Bos tauru	954	20	47.6	174114	12	AC073214	AC073214 Homo sapi
882	20.2	48.1	288211	12	AC181676	AC181676 Strongylo	955	20	47.6	177724	12	AC176849	AC176849 Strongylo
883	20.2	48.1	288211	12	AC181676	AC181676 Strongylo	956	20	47.6	181202	15	AF322012	AF322012 Bradyrhiz
884	20.2	48.1	294136	12	AC006901	AC006901 Caenorhab	957	20	47.6	180000	15	AF322012	AF322012 Homo sapi
885	20.2	48.1	309457	12	AC160280	AC160280 Bos tauru	958	20	47.6	182528	12	AC008048	AC008048 Homo sapi
886	20.2	48.1	326336	12	AC046137	AC046137 Homo sapi	959	20	47.6	184292	6	AC123848	AC123848 Mus muscu
887	20.2	48.1	328736	12	AC107298	AC107298 Homo sapi	960	20	47.6	184868	12	AC156769	AC156769 Rhinoph
888	20.2	48.1	336578	5	AP001745	AP001745 Homo sapi	961	20	47.6	185759	12	AC149612	AC149612 Ictalurus
889	20.2	48.1	340525	12	AC117847	AC117847 Rattus no	962	20	47.6	186114	5	AC021621	AC021621 Homo sapi
890	20.2	48.1	345109	12	AC098623	AC098623 Rattus no	963	20	47.6	186534	12	AC142482	AC142482 Rattus no
891	20.2	48.1	349061	15	NNA22491	BL62753 Neisseria	964	20	47.6	190295	6	AC124445	AC124445 Mus muscu
892	20.2	48.1	349648	15	BX957220	BX957220 Methanoco	965	20	47.6	191821	5	AC022616	AC022616 Homo sapi
893	20	47.6	527	2	CQ424285	CQ424285 Sequence	966	20	47.6	192632	6	AC164554	AC164554 Mus muscu
894	20	47.6	531	2	AX319681	AX319681 Sequence	967	20	47.6	192877	6	AC132286	AC132286 Mus muscu

c 968 20 47.6 193670 5 AC026261  
969 20 47.6 201361 11 BX927088  
970 20 47.6 204505 12 AC119344  
c 971 20 47.6 204836 12 AC119348  
c 972 20 47.6 205166 12 AC148169  
973 20 47.6 205338 12 AC141855  
974 20 47.6 205946 6 AC160135  
975 20 47.6 208408 12 AC179649  
976 20 47.6 209732 12 AC166400  
977 20 47.6 213907 12 CT573134  
978 20 47.6 215150 12 AC116438  
979 20 47.6 218565 12 AC097708  
980 20 47.6 219086 5 AC010147  
c 981 20 47.6 219548 6 AC165153  
c 982 20 47.6 223852 6 AC160113  
983 20 47.6 226256 6 AC125082  
984 20 47.6 227057 12 AC182137  
c 985 20 47.6 229214 12 CT030153  
c 986 20 47.6 229438 12 AC107408  
987 20 47.6 231559 12 AC111306  
988 20 47.6 233670 12 AC126740  
c 989 20 47.6 238933 12 AC096612  
c 990 20 47.6 239760 12 AC095113  
c 991 20 47.6 240425 6 CNS07YOT  
992 20 47.6 240447 6 CT009630  
993 20 47.6 241178 12 AC130508  
c 994 20 47.6 242832 12 AC129142  
c 995 20 47.6 245259 12 AC120898  
c 996 20 47.6 245808 12 AC114440  
c 997 20 47.6 246387 12 AC119756  
998 20 47.6 250175 12 AC160494  
c 999 20 47.6 252672 12 AC107498  
1000 20 47.6 253594 12 AC128893

ALIGNMENTS

RESULT 1  
BD233463 603 bp DNA linear PAT 17-JUL-2003  
LOCUS Human protein having hydrophobic domain and DNA encoding the same.  
DEFINITION BD233463  
ACCESSION BD233463.1 GI:33043233  
VERSION JP 2002519016-A/9.  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 603)  
AUTHORS Kato,S. and Kimura,T.  
TITLE Human protein having hydrophobic domain and DNA encoding the same  
JOURNAL Patent: JP 2002519016-A 9 02-JUL-2002;  
SAGAMI CHEMICAL RESEARCH CENTER,PROTEGENE INC  
COMMENT OS Homo sapiens (human)  
PN JP 2002519016-A/9  
PD 02-JUL-2002  
PF 18-JUN-1999 JP 2000557267  
PI SEISHI KATO,TOMOKO KIMURA  
PC C12N15/09,C07K14/47,C12N1/15,C12N1/19,C12N5/10,C12N15/00,C12N5/ PC  
00  
CC Human protein having hydrophobic domain and DNA encoding the  
CC same  
FH Key Location/Qualifiers  
FT source 1..603  
FT /organism='Homo sapiens (human)'.  
FT Location/Qualifiers  
1..603  
/organism='Homo sapiens'  
/mol\_type='genomic DNA'  
/db\_xref='taxon:9606'

ORIGIN  
Query Match 100.0%; Score 42; DB 2; Length 603;  
Best Local Similarity 100.0%; Pred. No. 8.9e-08;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GCCCAGCGTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
DB 550 GCCCAGCGTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 591  
RESULT 2  
BD209699 1212 bp DNA linear PAT 17-JUL-2003  
LOCUS Compositions isolated from skin cells and methods for their use.  
DEFINITION BD209699  
ACCESSION BD209699  
VERSION BD209699.1 GI:33019469  
KEYWORDS JP 2002512798-A/171.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1212)  
AUTHORS Strachan,L., Sleeman,M., Watson,J.D., Onrust,R., Kumble,A. and Murison,J.G.  
TITLE Compositions isolated from skin cells and methods for their use  
JOURNAL Patent: JP 2002512798-A 171 08-MAY-2002;  
GENESIS RESEARCH AND DEVELOPMENT CORP LTD  
COMMENT OS Homo sapiens (human)  
PN JP 2002512798-A/171  
PD 08-MAY-2002  
PF 29-APR-1999 JP 2000546009  
PR 09/069726,09-NOV-1998 US 09/188930 PI  
LORNA, STRACHAN, MATTHEW SLEEMAN, JAMES DOUGLAS WATSON, RENE PI  
ONRUST,  
PI ANAND KUMBLE, JAMES GREG MURISON  
PC C12N15/09,A61K38/00,A61P9/00,A61P17/00,A61P29/00,A61P31/18, PC  
A61P35/00,  
PC C07K14/47,C12N1/15,C12N1/19,C12N1/21,C12N5/10,C12P21/02,C12N15/ PC  
00,  
PC A61K37/02,C12N5/00  
CC Compositions isolated from skin cells and methods for their  
CC use.  
FH Key Location/Qualifiers  
FT source 1..1212  
FT /organism='Homo sapiens (human)'.  
FT Location/Qualifiers  
1..1212  
/organism='Homo sapiens'  
/mol\_type='genomic DNA'  
/db\_xref='taxon:9606'

ORIGIN  
Query Match 100.0%; Score 42; DB 2; Length 1212;  
Best Local Similarity 100.0%; Pred. No. 1.1e-07;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GCCCAGCGTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
DB 990 GCCCAGCGTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 1031  
RESULT 3  
AR341505 1212 bp DNA linear PAT 17-AUG-2003  
LOCUS AR341505  
DEFINITION Sequence 249 from patent US 6573095.  
ACCESSION AR341505  
VERSION AR341505.1 GI:33733640  
KEYWORDS Unknown.  
SOURCE Unknown.  
ORGANISM Unknown.

```

REFERENCE
AUTHORS      Unclassified.
TITLE        1 (bases 1 to 1212)
JOURNAL      Strachan,L.
             Polynucleotides isolated from skin cells
             Patent: US 6573095-A 249 03-JUN-2003;
             Genesis Research & Development Corporation Limited; Parnell;
             NZX;
FEATURES
source       Location/Qualifiers
             1..1212
             /organism="unknown"
             /mol_type="genomic DNA"
ORIGIN
Query Match 100.0%; Score 42; DB 2; Length 1212;
Best Local Similarity 100.0%; Pred. No. 1.1e-07;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCCACGCTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42
   |||||||
Db 990 GCCACGCTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1031
   |||||||

RESULT 4
LOCUS      AX078375 1619 bp DNA linear PAT 22-FEB-2001
DEFINITION Sequence 43 from Patent WO0107612.
ACCESSION  AX078375
VERSION     AX078375.1 GI:13158044
KEYWORDS   Homo sapiens (human)
SOURCE     Homo sapiens
ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
           Homnidae; Homo.
REFERENCE
AUTHORS    Au-Young,J., Bandman,O., Tang,Y.T., Yue,H., Azimzai,Y., Burford,N.,
           Baughn,M.R., Lu,D.A., Hillman,J.L., Patterson,C. and Lal,P.
TITLE      Receptors and associated proteins
JOURNAL    Patent: WO 0107612-A 43 01-FEB-2001;
           Incyte Genomics, Inc. (US)
FEATURES
source     Location/Qualifiers
             1..1619
             /organism="Homo sapiens"
             /mol_type="unassigned DNA"
             /db_xref="taxon:9606"
             /note="Incyte ID No: 2681738CB1"
ORIGIN
Query Match 100.0%; Score 42; DB 2; Length 1619;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCCACGCTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42
   |||||||
Db 1137 GCCACGCTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1178
   |||||||

RESULT 5
LOCUS      BD233473 1718 bp DNA linear PAT 17-JUL-2003
DEFINITION Human protein having hydrophobic domain and DNA encoding the same.
ACCESSION  BD233473
VERSION     BD233473.1 GI:33043243
KEYWORDS   JP 2002519016-A/19.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
           Homnidae; Homo.
REFERENCE
AUTHORS    Kato,S. and Kimura,T.
TITLE      Human protein having hydrophobic domain and DNA encoding the same
JOURNAL    Patent: JP 2002519016-A 19 02-JUL-2002;

```

---

```

COMMENT      SAGAMI CHEMICAL RESEARCH CENTER, PROTEGENE INC
OS           Homo sapiens (human)
PN           JP 2002519016-A/19
PD           02-JUL-2002
PF           18-JUN-1999 JP 2000557267
PI           SEISHI KATO,TOMOKO KIMURA
PC           C12N15/09,C07K14/47,C12N1/15,C12N1/19,C12N5/10,C12N15/00,C12N5/ PC
00
CC           Human protein having hydrophobic domain and DNA encoding the
FH           same
FT           Key source 1..1718
           Location/Qualifiers
           /organism="Homo sapiens (human)"
           /mol_type="genomic DNA"
           /db_xref="taxon:9606"
FEATURES
source       Location/Qualifiers
             1..1718
             /organism="Homo sapiens"
             /mol_type="genomic DNA"
             /db_xref="taxon:9606"
ORIGIN
Query Match 100.0%; Score 42; DB 2; Length 1718;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCCACGCTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42
   |||||||
Db 561 GCCACGCTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 602
   |||||||

RESULT 6
LOCUS      AK172760 1826 bp mRNA linear PRI 07-MAY-2004
DEFINITION Homo sapiens cDNA FLJ23921 fis, clone COL02043, highly similar to
           Homo sapiens retinoic acid induced 3 (RAI3).
ACCESSION  AK172760
VERSION     AK172760.1 GI:47077732
KEYWORDS   oligo capping; fis (full insert sequence).
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
           Homnidae; Homo.
REFERENCE
AUTHORS     Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,
           Okamoto,S., Okitani,K., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
           Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
TITLE      NEDO human cDNA sequencing project
JOURNAL    Unpublished
REFERENCE   2 (bases 1 to 1826)
AUTHORS     Sugano,S. and Suzuki,Y.
JOURNAL    Direct Submission
           Submitted (22-APR-2004) Sumio Sugano, Institute of Medical Science,
           University of Tokyo, Laboratory of Genome Structure, Human Genome
           Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
           (E-mail:flcdna@mail.ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
           Fax:81-3-5449-5416)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
           Economy, Trade and Industry of Japan; cDNA full insert sequencing;
           Research Association for Biotechnology; cDNA library construction;
           5'- & 3'-end one pass sequencing; Department of Virology and Human
           Genome Center, Institute of Medical Science, University of Tokyo
           (partly supported by Science and Technology Agency).
FEATURES
source     Location/Qualifiers
             1..1826
             /organism="Homo sapiens"
             /mol_type="mRNA"
             /db_xref="taxon:9606"
             /clone="COL02043"
             /tissue_type="colon"
             /clone_lib="COL"
             /note="cloning vector pME18SFL3"
ORIGIN

```

```

Query Match      100.0%; Score 42; DB 5; Length 1826;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCCAGCGTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42
Db 638 GCCCAGCGTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 679

RESULT 7
LOCUS CQ723177 2290 bp DNA linear PAT 03-FEB-2004
DEFINITION Sequence 9111 from Patent WO02068579.
ACCESSION CQ723177
VERSION CQ723177.1 GI:42284034
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1
AUTHORS Venter,C.J., Adams,M.C., Li,P.W. and Myers,E.W.
TITLE Kites, such as nucleic acid arrays, comprising a majority of
humanexons or transcripts, for detecting expression and other uses
thereof
JOURNAL Patent: WO 02068579-A 9111 06-SEP-2002;
PE Corporation (NY) (US)
FEATURES
source
1..2290
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match      100.0%; Score 42; DB 2; Length 2290;
Best Local Similarity 100.0%; Pred. No. 1.4e-07;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCCAGCGTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42
Db 1117 GCCCAGCGTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1158

RESULT 8
LOCUS BC003665 2296 bp mRNA linear PRI 24-NOV-2004
DEFINITION Homo sapiens G protein-coupled receptor, family C, group 5, member
A, mRNA (CDNA clone MGC:923 IMAGE:2988011), complete cds.
ACCESSION BC003665
VERSION BC003665.2 GI:33872669
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 2296)
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Haieh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Iqbal,N.A., Peters,G.J.,
Abramson,R.D., Mullen,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Vitaloni,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,

```

```

Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butcherfield,Y.S., Krzywinski,M.I., Skalska,U., Smalilus,D.E.,
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 2296)
Director MGC Project.
Direct Submission
Submitted (26-FEB-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
On Aug 19, 2003 this sequence version replaced gi:13177795.
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbio.org
contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettman, Anuradha
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 3 Row: 1 Column: 11
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 12056470.
FEATURES
Location/Qualifiers
1..2296
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:923 IMAGE:2988011"
/tissue type="Colon, adenocarcinoma"
/clone_lib="NIH MGC 15"
/lab_host="DH10B-R"
/note="vector: POTB7"
1..2296
/gene="GPCR5A"
/note="synonyms: RAIG1, GPCR5A, RA13"
/db_xref="GeneID:9052"
/db_xref="MIM:604138"
95..1168
/gene="GPCR5A"
/codon_start=1
/product="G protein-coupled receptor, family C, group 5,
member A"
/protein_id="AAH03665.1"
/db_xref="GI:13177796"
/db_xref="GeneID:9052"
/db_xref="MIM:604138"
/translation="MATTTPDGCNRLKSKYVRLCDKAEAMGIVLETVATAGVWTSVA
FMLTPILVCKVQDSNRRKMLPTQFLVLGVLGFLTFATF IGLDGSIGTPRFLFG
ILFSICSCULAHAVSLTKVGRKPLSLVLGLAVGFSIAVDVIAEYIVLTQNRFT
NVNVSLSAPLRNEDFVLLTYVYFLMALTFILMSSFTFCGFTGKWRGHAIYLTML
LSIAIWWAITLMLPDRRDDTIILSLAANGWVFLAYVSPFVLLTKQRPMD
YVPDFAFCPLQVKKSYGVENRAYSBEITQGFETGDTLYAPYSTHFLQNPQPPQKE
FSIPRAHWPSPKYKDYEVKKEGS"
ORIGIN
Query Match      100.0%; Score 42; DB 5; Length 2296;
Best Local Similarity 100.0%; Pred. No. 1.4e-07;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCCAGCGTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42
Db 1112 GCCCAGCGTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1153

```



RESULT 9  
AF506289  
LOCUS Homo sapiens orphan G protein-coupling receptor PEIG-1 mRNA, PRI 15-JUL-2002 complete cds.  
ACCESSION AF506289  
VERSION 1 GI:21779962  
KEYWORDS Homo sapiens (human)  
SOURCE ORGANISM  
REFERENCE 1 (bases 1 to 2297)  
AUTHORS Cafferata,E.G., Gonzalez-Guerrico,A.M., Costanzo,R., Pivetta,O.H. and Santa-Coloma,T.A.  
TITLE Identification by differential display of a mRNA specifically induced by 12-O-tetradecanoylphorbol-13-acetate (TPA) in T84 human colon carcinoma cells  
JOURNAL Cell. Mol. Biol. 42 (5), 797-804 (1996)  
PUBMED 8832110  
AUTHORS 2 (bases 1 to 2297)  
Cafferata,E.G., Gonzalez-Guerrico,A.M., Costanzo,R., Pivetta,O.H. and Santa-Coloma,T.A.  
TITLE Direct Submission  
JOURNAL Submitted (24-APR-2002) Laboratorio de Biologia Celular y Molecular, Instituto de Investigaciones Bioquimicas Fundacion Campomar, Patricias Argentinas 435, Buenos Aires 1405, Argentina Location/Qualifiers  
FEATURES  
source  
1..2297  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/chromosomes="12"  
/map="12p13-p12.3"  
/sex="male"  
/cell\_line="T84"  
/tissue\_type="colorectal carcinoma"  
/note="derived from dbEST BE519991"  
108..1181  
/note="belongs to the type 3 G protein-coupling receptor family; characterized by a 7-transmembrane domain motif; phorbol ester induced protein-1"  
/codon\_start=1  
/product="orphan G protein-coupling receptor PEIG-1"  
/protein\_id="AAM77594.1"  
/db\_xref="GI:21779963"  
FMLTPIILVCKVQDSNRKMLPTQFLFLGLVIGLTFGLTFAFIIGLDGSGTPRFLFG ILFSICFCLLAHVASLTKLVGRKPLSLVILGLAVGFSLVQDVIAIEYIVLTWRT NVNVSLSAPRRNEDFVLLLTYYVLFMLATFLMSSFTFCGSFTGWRKHGAHYLTML LSIATWAVITLLMLPDFDRWDDTILSSALAANGVFLIAYVPEFWLLTKQRNPM YPEDAFCKPQLVKGKGVENRAYSQEIIQTGFEETGDTLVAPYSTHFLQNLQNPQKE PSIPRAHAWSPYKDYEVKKEGS"  
ORIGIN  
Query Match 100.0%; Score 42; DB 5; Length 2297;  
Best Local Similarity 100.0%; Pred. No. 1.4e-07;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GCCCAGCGTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
|||||  
DB 1125 GCCCAGCGTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1166  
RESULT 10  
AX930411  
LOCUS AX930411  
DEFINITION Sequence 2 from Patent WO03087832.  
ACCESSION AX930411  
VERSION AX930411.1 GI:40312209

KEYWORDS Homo sapiens (human)  
SOURCE ORGANISM  
REFERENCE 1  
AUTHORS Terrett,J.A.  
TITLE Diagnosis of carcinoma using raigl polypeptides  
JOURNAL Patent: WO 03087832-A 2 23-OCT-2003;  
Oxford GlycoSciences (UK) Limited (GB)  
FEATURES  
source  
1..2302  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Query Match 100.0%; Score 42; DB 2; Length 2302;  
Best Local Similarity 100.0%; Pred. No. 1.4e-07;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GCCCAGCGTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
|||||  
DB 1117 GCCCAGCGTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1158  
RESULT 11  
AF095448  
LOCUS Homo sapiens putative G protein-coupled receptor (RAIG1) mRNA, PRI 29-DEC-1998  
DEFINITION complete cds.  
ACCESSION AF095448  
VERSION AF095448.1 GI:4063889  
KEYWORDS  
SOURCE ORGANISM  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 2302)  
AUTHORS Cheng,Y. and Lotan,R.  
TITLE Molecular cloning and characterization of a novel retinoic acid-inducible gene that encodes a putative G protein-coupled receptor  
JOURNAL J. Biol. Chem. 273 (52), 35008-35015 (1998)  
PUBMED 9857033  
REFERENCE 2 (bases 1 to 2302)  
AUTHORS Cheng,Y. and Lotan,R.  
TITLE Direct Submission  
JOURNAL Submitted (27-SEP-1998) Tumor Biology, The University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX 77030, USA  
FEATURES  
source  
1..2302  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/chromosome="12"  
/map="12p12.3-p13"  
1..2302  
/gene="RAIG1"  
/note="induced by retinoic acid"  
100..1173  
/gene="RAIG1"  
/codon\_start=1  
/product="putative G protein-coupled receptor"  
/protein\_id="AAC98506.1"  
/db\_xref="GI:4063889"  
FMLTPIILVCKVQDSNRKMLPTQFLFLGLVIGLTFGLTFAFIIGLDGSGTPRFLFG ILFSICFCLLAHVASLTKLVGRKPLSLVILGLAVGFSLVQDVIAIEYIVLTWRT NVNVSLSAPRRNEDFVLLLTYYVLFMLATFLMSSFTFCGSFTGWRKHGAHYLTML

LSIAIWAWITLLMLPDRDWDITLSSALAANGWFLLAYVSPFWLLTKQRNPM  
YPVEDAFCKPOLVKSYGVENRAYSQEITQGFETGDTLYAPYSTHFLQONQPQKE  
FSIPRAHAWPSYKDYEVKKEGS"

ORIGIN

Query Match 100.0%; Score 42; DB 5; Length 2302;  
Best Local Similarity 100.0%; Pred. No. 1.4e-07;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCACGCTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42

Db 1117 GCCCACGCTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1158

RESULT 12

BD156680 2446 bp DNA linear PAT 17-JAN-2003  
LOCUS Primer for synthesizing full-length cDNA and use thereof.  
DEFINITION  
ACCESSION BD156680  
VERSION BD156680.1 GI:27862438  
KEYWORDS JP 2002191363-A/11523.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominiidae; Homo.

1 (bases 1 to 2446)

Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,  
Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.  
Primer for synthesizing full-length cDNA and use thereof  
Patent: JP 2002191363-A 11523 09-JUL-2002;

COMMENT

HELIIX RESEARCH INSTITUTE  
OS Homo sapiens (human)  
PN JP 2002191363-A/11523  
PD 09-JUL-2002  
PF 28-JUL-2000 JP 2000280990  
PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU  
PI SAITO,  
PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,  
PI KEIICHI NAGAI, TETSUJI OTSUKI  
PC  
C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC  
10,  
PC C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC  
Primer for synthesizing full-length cDNA and use thereof FH Key  
Location/Qualifiers  
FT CDS (254). .(1324).

FEATURES

source  
1..2446  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"

ORIGIN

Query Match 100.0%; Score 42; DB 2; Length 2446;  
Best Local Similarity 100.0%; Pred. No. 1.4e-07;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCACGCTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42

Db 1271 GCCCACGCTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1312

RESULT 13

AX877483 2446 bp DNA linear PAT 17-DEC-2003  
LOCUS Sequence 12388 from Patent EP1074617.  
DEFINITION  
ACCESSION AX877483  
VERSION AX877483.1 GI:40032219  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE

AUTHORS

TITLE  
JOURNAL

FEATURES

source

1..2446  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
254..1327  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAE89746.1"  
/db\_xref="GI:40032220"  
/translation="MATTPDGCRCGLKSKYVRLCDKAEANGIVLETVATAGVTSVA  
FMLPLILVCKVQDSNRKMLPTQFLFLVLGLFGLTLPAPILGLDGSSTGPTFFLEF  
ILFICFCLLAHAVSLTKVGRKPLSLVILGLAVGSLVQDVIALEYIVLTWRT  
NNVFSELSAPRNEFDVLLTYVFLMALTFMSSFTFCGSGTCKRHGAHYITML  
LSIAIWAWITLLMLPDRDWDITLSSALAANGWFLLAYVSPFWLLTKQRNPM  
YPVEDAFCKPOLVKSYGVENRAYSQEITQGFETGDTLYAPYSTHFLQONQPQKE  
FSIPRAHAWPSYKDYEVKKEGS"

ORIGIN

Query Match 100.0%; Score 42; DB 2; Length 2446;  
Best Local Similarity 100.0%; Pred. No. 1.4e-07;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCACGCTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42

Db 1271 GCCCACGCTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1312

RESULT 14

AK001761 2446 bp mRNA linear PRI 20-JAN-2006  
LOCUS Homo sapiens CDNA FLJ10899 fis, clone NT2RP5003506.  
DEFINITION  
ACCESSION AK001761  
VERSION AK001761.1 GI:7023229  
KEYWORDS oligo capping; fis (full insert sequence).  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominiidae; Homo.

REFERENCE

AUTHORS

Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R.,  
Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H.,  
Sekine, M., Obayashi, M., Nishi, T., Shibahara, T., Tanaka, T.,  
Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y.,  
Nagahari, K., Murakami, K., Yasuda, T., Iwayanagi, T., Wagatsuma, M.,  
Shiratori, A., Sudo, H., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H.,  
Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T.,  
Kikkawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K.,  
Tanikawa, M., Yamazaki, M., Ninomiya, K., Ishibashi, I., Yamashita, H.,  
Murakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M.,  
Hiraoka, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S.,  
Yosida, M., Hotuta, T., Kusano, J., Kanehori, K., Takahashi-Fujii, A.,  
Hara, H., Tanase, T.O., Nomura, Y., Togiya, S., Komai, F., Hara, R.,  
Takeuchi, K., Arita, M., Imose, N., Musashino, K., Yuuki, H., Oshima, A.,  
Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsumawa, H., Ichihara, T.,  
Shiohata, N., Sano, S., Moriya, S., Momiyama, H., Satoh, N., Takami, S.,  
Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H.,  
Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T.,  
Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K.,  
Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M.,  
Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yamada, K., Fujii, Y.,  
Ozaki, K., Hirao, M., Ohmori, Y., Kawabata, A., Hikiji, T., Kobatake, N.,  
Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Kawakami, T.,  
Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsumura, K.,

Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T., Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J., Sato, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K., Nagase, T., Nomura, N., Kikuchi, H., Masuho, Y., Yamashita, R., Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S. Complete sequencing and characterization of 21,243 full-length human cDNAs  
Nat. Genet. 36 (1), 40-45 (2004)

2 Isogai, T., Ota, T., Hayaishi, K., Sugiyama, T., Otsuki, T., Suzuki, Y., Nishikawa, T., Nagai, K., Sugano, S., Ishibashi, T., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M., Hiraoka, S., Ishii, S., Kawai, Y., Saito, K., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagahori, K., Masuho, Y. and Kanehori, K.  
NEDO human cDNA sequencing project  
Unpublished  
3 (bases 1 to 2446)  
Isogai, T. and Otsuki, T.  
Direct Submission  
Submitted (16-FEB-2000) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: flij-cdna@nifty.com, Tel: 81-438-52-3975, Fax: 81-438-52-3986)  
NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing and clone selection; Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.  
Location/Qualifiers  
1..2446  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="NT2RP5003506"  
/cell\_line="NT2"  
/cell\_type="teratocarcinoma"  
/clone\_lib="NT2RPS"  
/note="cloning vector: pME18SFL3  
mRNA from NT2 neuronal precursor cells after 2-weeks retinoic acid (RA) induction."  
254..1327  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="BAA91890.1"  
/db\_xref="GI:7023230"  
/translation="MATTVPDGRNGLSKYYRLCDKAEANGIVLETVATAGVVTSSA  
FMLTILVCKVQDSNRKMLPTQFLLLGLVIGLTFATIGLDGSGTPTRPFLFG  
ILFSICFSLLAHVSFLKLVGRKPLSLVILGLVGFSLVQDVIALEYLVLTNRT  
NVNFSLSAPRNEFDVLLTYLVFLMALFLMSFTFCGSFTGWRKHGHIYLTML  
LSIAIWAWITLLMLPDDRWDITLSSALAANGWVLLAYVSPFWLLTKQRNPMD  
YVPEDAFKPQLVKKSYGVENRAYSQEITQGFETGDTLYAPYSTHFLQLNQPPQKE  
FSIPRAHWPSPKDYEVKKEGS"

CDs

Query Match 100.0%; Score 42; DB 5; Length 2446;  
Best Local Similarity 100.0%; Pred. No. 1.4e-07;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
|||||  
Db 1271 GCCCAGCGTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1312

RESULT 15  
LOCUS CQ981495  
DEFINITION Sequence 350 from Patent EP1498424.  
ACCESSION CQ981495  
VERSION CQ981495.1 GI:58190785  
KEYWORDS Homo sapiens (human)

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
AUTHORS Rosenthal, A., Hermann, K., Heiden, E., Pilarsky, C., Bruemendorf, T., Staub, E., Roepcke, S., Mennerich, D., Kinnemann, H. and Li, X.  
TITLE Human nucleic acid sequences from lung tumours  
JOURNAL Patent: EP 1498424-A 350 13-JAN-2005; Heiden, Esmeralda (DE); Hinzmann, Bernd (DE); Hermann, Klaus (DE); Rosenthal, Andre (DE)

FEATURES  
source  
1..2456  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

ORIGIN  
Query Match 100.0%; Score 42; DB 2; Length 2456;  
Best Local Similarity 100.0%; Pred. No. 1.4e-07;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
|||||  
Db 1271 GCCCAGCGTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1312

RESULT 16  
LOCUS DD210040  
DEFINITION Methods of Diagnosis of Cancer, Compositions and Methods of Screening for Modulators of Cancer.

ACCESSION DD210040  
VERSION DD210040.1 GI:85654022  
KEYWORDS JP 2005518782-A/34.  
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
AUTHORS Zlotnik, A., Mack, D.H., Agiz, N., Gish, K.C., Hebeji, P.A., Wilson, K.E. and Afar, D.  
TITLE Methods of Diagnosis of Cancer, Compositions and Methods of Screening for Modulators of Cancer  
JOURNAL Patent: JP 2005518782-A 34 30-JUN-2005; PROTEIN DESIGN LABS INC

COMMENT  
OS Homo sapiens  
PN JP 2005518782-A/34  
PD 30-JUN-2005  
PF 17-SEP-2002 JP 2003523912  
PR 12-APR-2002 US 60/372246, 08-FEB-2002 US 60/355257, PR 08-FEB-2002 US 60/355145, 13-NOV-2001 US 60/350666, PR 20-SEP-2001 US 60/323887, 17-SEP-2001 US 60/323469 PI albert  
Zlotnik, david h mack, natasha agiz, kurt c gish, peter a pi hebeji, PI keith e wilson, daniel afar

CC  
FH  
Location/Qualifiers  
1..2456  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

ORIGIN  
Query Match 100.0%; Score 42; DB 2; Length 2456;  
Best Local Similarity 100.0%; Pred. No. 1.4e-07;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
|||||  
Db 1271 GCCCAGCGTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1312

```

RESULT 17
AX549168
LOCUS
DEFINITION Sequence 453 from Patent WO02061087.
ACCESSION AX549168
VERSION AX549168.1 GI:25813894
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE
AUTHORS 1
TITLE Antigenic peptides, such as for G protein-coupled receptors
(GPCRs), antibodies thereto, and systems for identifying such
antigenic peptides
JOURNAL Patent: WO 02061087-A 453 08-AUG-2002;
Lifespan Biosciences, Inc. (US)
FEATURES
source
1..2456
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 100.0%; Score 42; DB 2; Length 2456;
Best Local Similarity 100.0%; Pred. No. 1.4e-07;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCACGGTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42
|||||
Db 1271 GCCACGGTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1312

RESULT 18
AK122672
LOCUS
DEFINITION Homo sapiens CDNA FLJ16117 fis, clone ASTRO2003632, highly similar
to Homo sapiens putative G protein-coupled receptor (RAIG1) mRNA.
ACCESSION AK122672
VERSION AK122672.1 GI:34527861
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE
AUTHORS 1
TITLE Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R.,
Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H.,
Sekine, M., Obayashi, M., Nishi, T., Shibahara, T., Tanaka, H.,
Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y.,
Negahari, K., Murakami, K., Yasuda, T., Iwayanagi, T., Wagatsuma, M.,
Shiratori, A., Sudoh, H., Hosoiri, T., Kaku, K., Yokoi, T., Furuya, T.,
Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Kodaira, H., Kondo, H.,
Kikkawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K.,
Tanikawa, M., Yamazaki, M., Ninomiya, K., Ishibashi, T., Yamashita, H.,
Murakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M.,
Hiraoaka, S., Chiba, Y., Ishida, S., Oho, Y., Takiguchi, S., Watanabe, S.,
Yosida, M., Hotta, T., Kusano, J., Kanehori, K., Takahashi-Fujii, A.,
Hara, H., Tanase, T. O., Nomura, Y., Togiya, S., Komai, F., Hara, R.,
Takeuchi, K., Arita, M., Imose, N., Musashino, K., Yuuki, H., Oshima, A.,
Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T.,
Shiohata, N., Sano, S., Moriya, S., Momiyama, H., Satoh, N., Takami, S.,
Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H.,
Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T.,
Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K.,
Kunagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, K.,
Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yamada, K., Fujii, Y.,
Ozaki, K., Hirao, M., Ohmori, Y., Kawabata, A., Hikiji, T., Kobatake, N.,
Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Kawakami, T.,

```

```

Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsumura, K.,
Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T.,
Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J.,
Sato, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K.,
Nagase, T., Nomura, N., Kikuchi, H., Masuo, Y., Yamashita, R.,
Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S.
Complete sequencing and characterization of 21,243 full-length
human cDNAs
Nat. Genet. 36 (1), 40-45 (2004)
14702039
2
Tanigami, A., Fujiwara, T., Shibahara, T., Goto, Y., Hirao, M.,
Shimizu, F., Wakebe, H., Oho, T., Hishigaki, H., Watanabe, T., Ozaki, K.,
Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S.,
Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T.,
Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M.,
Kikuchi, H., Kanda, K., Wagatsuma, M., Murakawa, K., Kanehori, K.,
Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y.,
Sugano, S., Negahari, K., Masuo, Y., Nagai, K. and Isogai, T.
NEDO human cDNA sequencing project
Unpublished
3 (bases 1 to 3057)
Isogai, T. and Yamamoto, J.
Direct Submission
Submitted (15-JUL-2003) Takao Isogai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail: flj-cdna@nifty.com, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: Reverse Proteomics Research Institute, HRI and
RAB.
FEATURES
source
1..3057
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="ASTRO2003632"
/cell_type="normal astrocytes (NHA5732)"
/clone_lib="ASTRO2"
/note="cloning vector: pWE18SFL3
primary culture, normal astrocytes"
ORIGIN
Query Match 100.0%; Score 42; DB 5; Length 3057;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCACGGTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42
|||||
Db 1908 GCCACGGTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1949

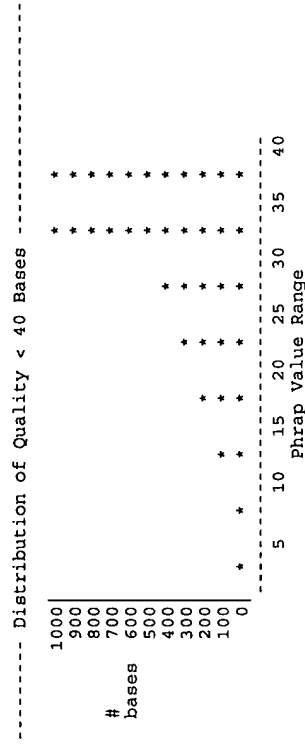
RESULT 19
AX188348
LOCUS
DEFINITION Sequence 4043 from Patent WO0142467.
ACCESSION AX188348
VERSION AX188348.1 GI:15139821
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE
AUTHORS 1
TITLE Schlegel, R., Deeds, J., Berger, A. and Zhao, X.
Genes, compositions, kits, and methods for identification,
assessment, prevention, and therapy of cervical cancer

```

JOURNAL Patent: WO 0142467-A 4043 14-JUN-2001;  
Millennium Predictive Medicine, Inc. (US)  
FEATURES Location/Qualifiers  
source 1..3371  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Query Match 100.0%; Score 42; DB 2; Length 3371;  
Best Local Similarity 100.0%; Pred. NO. 1.6e-07;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GCCACGCTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
|||||  
Db 1184 GCCACGCTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1225  
|||||  
RESULT 20  
CQ894732 6730 bp DNA linear PAT 05-NOV-2004  
DEFINITION Sequence 42 from Patent EP1471075.  
ACCESSION CQ894732  
VERSION CQ894732.1 GI:55467481  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
REFERENCE 1 Rosenthal,A.D., Pilarsky,C., Dahl,E., Specht,T., Bruemendorf,T.,  
Lichtner,R., Staub,E., Roepcke,S. and Li,X.I.  
Human nucleic acid sequences expressed in pancreatic carcinomas  
Patent: EP 1471075-A 42 27-OCT-2004;  
Hinzmann, Bernd (DE); Rosenthal, Andre (DE); Pilarsky, Christian  
(DE); Dahl, Edgar (DE); Specht, Thomas (DE); Lichtner, Rosemarie  
(DE)  
FEATURES Location/Qualifiers  
source 1..6730  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Query Match 100.0%; Score 42; DB 2; Length 6730;  
Best Local Similarity 100.0%; Pred. NO. 2.1e-07;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GCCACGCTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
|||||  
Db 1271 GCCACGCTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1312  
|||||  
RESULT 21  
AC007688/c 161577 bp DNA linear PRI 30-AUG-2002  
LOCUS AC007688 Human BAC Library) complete sequence.  
DEFINITION Homo sapiens 12 BAC RPC111-392P7 (Roswell Park Cancer Institute  
Human BAC Library) complete sequence.  
ACCESSION AC007688  
VERSION AC007688.15 GI:5815499  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
REFERENCE 1 (bases 1 to 161577)  
AUTHORS Muzny,D., Arenson,A.D., Bouck,J., Bunac,C., Chen,Z., Ding,Y.,  
Dugan,S., Durbin,J., Forcum,J., Garcia,C., Correll,J.H.,  
Gorrell,L.B., Hernandez,J., Issar,A., Jackson,L., Kneitz,S.,  
Kondejewski,N., Lau,S., Leal,B., Lee,E., Lichtarge,O., Liu,W.,  
Logan,O., Lu,J., Marondel,I., Martinez,C., Merscher,S., Miller,A.,  
Montgomery,K., Oswal,G., Pampell,L.R., Parish,B.J., Perez,L.,  
Rashid,N.D., Rives,C., Scherer,S.E., Shen,H., Shim,C., Simon,M.,  
Vo,Q., Williamson,A., Worley,K.C., Shang,A.M., Yang,R., Yu,W.,  
Zhou,X., Kucherlapati,R., Nelson,D. and Gibbs,R.A.  
Direct Submission  
Unpublished  
2 (bases 1 to 161577)  
Worley,K.C.  
Direct Submission  
Submitted (01-JUN-1999) Molecular and Human Genetics, Baylor  
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 161577)  
Worley,K.C.  
Direct Submission  
Submitted (01-SEP-1999) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
4 (bases 1 to 161577)  
Worley,K.C.  
Direct Submission  
Submitted (16-MAY-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
5 (bases 1 to 161577)  
Worley,K.C.  
Direct Submission  
Submitted (30-AUG-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Sep 1, 1999 this sequence version replaced gi:5757565.  
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email  
[gc-help@bcm.tmc.edu](mailto:gc-help@bcm.tmc.edu)  
CLONE LENGTH: This sequence does not necessarily represent the  
entire insert of this clone. Overlapping regions of clones are only  
sequenced and submitted once, so the sequence for the remainder of  
the insert may be found in the record for the adjacent clones.  
Overlapping clones are noted at the beginning and end of the  
Features listing.  
ANNOTATION OF FEATURES:  
STSs are identified using ePCR (Genome Res. 7:541-550) searches  
of a local database that includes entries from dbSTS, GDB, and  
local mapping efforts.  
Repeats are identified using RepeatMasker (A. Smit and P. Green,  
unpublished.) for Human and Mouse sequences.  
Genes and Region of sequence similarity are identified by BLAST  
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the  
EST and cDNA sequences. Genes demonstrate at least two exons  
flanked by consensus splice sites that maintained sequence  
continuity across the splice junctions. Sequences that are not  
identical matches are annotated as similar.  
SEQUENCING READ COVERAGE: Sequencing is completed to a minimum  
standard of double strand coverage with a minimum of 2 clones and 2  
reads with no ambiguities or 2 chemistries with a minimum of 2  
clones and 3 reads with no ambiguities. If the sequence quality for  
a region does not meet this standard, it will be indicated in the  
annotation as Low Coverage.  
QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality  
standards - estimated error rate less than 1 per 10,000 bases.  
Reports of lowest quality individual bases and measures of base  
quality are listed below. Description of the metrics can be found  
at URL:  
<http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.  
QUALSTAT-REPORT-----  
----- Summary Statistics -----  
Contig length: 161577  
Phrap values in estimate: 160751

Average error rate (BCM-Phrap estimate): 0.000163681  
Fraction of Phrap values less than 40: 0.0376047  
Number of consensus changing edits: 30  
Number of N's in consensus: 0

```
----- Consensus changing edits -----
Position      Original Context      Edited Context
7033          acctcgccgt(n)ccgcccctt acctcgccgt(n)ccgcccctt
47567         aaaaaaanaa(n)ggaataaat aaaaaaanaa(n)ggaataaat
51135         aaagaagaag(n)aaagaaggaa aaagaagaag(n)aaagaaggaa
75582         aaaaaaanaa(n)aaaggatgct aaaaaaanaa(n)aaaggatgct
75585         aaaaaaanaa(n)ggaatgtcgc aaaaaaanaa(n)ggaatgtcgc
75587         ctaaaagcaga(n)taagatctta ttttaaatag(n)gctttgcttt
75752         ttttaaatag(n)gctttgcttt ttttaaatag(n)gctttgcttt
84017         gggaggggaag(n)aaaggaaggag gggaggggaag(n)aaaggaaggag
85227         tttgtttttt(n)tggttttttt tttgtttttt(n)tggttttttt
99681         ggaagtccaag(n)atgcagtgag ggaagtccaag(n)atgcagtgag
111307        aatctcttat(n)ccgaattcca aatctcttat(n)ccgaattcca
111374        actagtatac(n)atentttttt actagtatac(n)atentttttt
111378        gtatacnatc(n)tttttttttt gtatacnatc(n)tttttttttt
112621        aaaaaaanaa(n)ccatctctaga aaaaaaanaa(n)ccatctctaga
115812        attcaccttc(n)tttttttttt attcaccttc(n)tttttttttt
137207        ttgcaggcac(n)cgccaccacg ttgcaggcac(n)cgccaccacg
137218        cgccaccacg(n)ctggctaaat cgccaccacg(n)ctggctaaat
145113        gcaagtgaag(n)natgtagaat gcaagtgaag(n)natgtagaat
145114        caagtgaan(n)atgtagaata caagtgaan(n)atgtagaata
145232        ggcaccgggt(n)ntcancncag ggcaccgggt(n)ntcancncag
145233        ggcaccgggt(n)tcacnccagt ggcaccgggt(n)tcacnccagt
145238        ggtgnntcac(n)ccagtaatcc ggtgnntcac(n)ccagtaatcc
145321        gcaacatggt(n)nncccatcgc gcaacatggt(n)nncccatcgc
145322        caacatggtt(n)nncccatctc caacatggtt(n)nncccatctc
145323        aacatggttn(n)cccctctctc aacatggttn(n)cccctctctc
145324        acatggtttn(n)cccctctctc acatggtttn(n)cccctctctc
145363        gtccaccagg(n)gtgtggcgt  gtccaccagg(n)gtgtggcgt
145377        gtggcgcgta(n)ttgtagtcct gtggcgcgta(n)ttgtagtcct
145397        tagctacttg(n)gagggaggat tagctacttg(n)gagggaggat
145505        aaaaaaaaaa(n)gaaaaaaaat aaaaaaaaaa(n)gaaaaaaaat
```



Version: 1.01 qxfo.  
Location/Qualifiers  
1..161577  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="12"  
/clone="RP11-392P7"  
complement(3..65)  
/rpt\_family="MER4B"  
161..471  
/rpt\_family="AluXs"  
complement(607..702)  
/rpt\_family="MIR"

```
repeat_region 703..838  
/rpt_family="FLAM C"  
repeat_region complement(839..916)  
/rpt_family="MIR"  
repeat_region complement(1205..1377)  
/rpt_family="MER104"  
STS 1388..1525  
/standard_name="WIAF-759-STS"  
/db_xref="dbSTS:65358"  
repeat_region 1683..1971  
/rpt_family="AluJo"  
repeat_region 1974..2261  
/rpt_family="Aluub"  
repeat_region 2486..2593  
/rpt_family="MIR"  
misc_feature 2610..3214  
/note="Region: Unigene cluster similar to AA056332 and  
AI074576"  
STS 2712..2823  
/standard_name="SHGC-44583"  
/db_xref="dbSTS:48451"  
repeat_region 3185..3491  
/rpt_family="AluSg"  
repeat_region complement(6397..6537)  
/rpt_family="L2"  
repeat_region 6967..7158  
/rpt_family="(CCCCG)n"  
repeat_region 7094..7240  
/rpt_family="(CCG)n"  
repeat_region complement(7770..8250)  
/rpt_family="MER44B"  
repeat_region 8831..9073  
/rpt_family="MER102"  
repeat_region complement(9075..9303)  
/rpt_family="L2"  
repeat_region 9671..10135  
/rpt_family="LTR33A"
```

Query Match 100.0%; Score 42; DB 5; Length 161577;  
Best Local Similarity 100.0%; Pred. No. 6.2e-07;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 GCCCAGCGTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
DB 94911 GCCCAGCGTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 94870
```

RESULT 22  
BD150346 680 bp DNA linear PAT 17-JAN-2003  
LOCUS Primer for synthesizing full-length cDNA and use thereof.  
DEFINITION BD150346  
ACCESSION BD150346  
VERSION BD150346.1 GI:27856104  
KEYWORDS JP 2002191363-A/5189.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
REFERENCE 1 (bases 1 to 680)  
AUTHORS Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,  
Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.  
TITLE Primer for synthesizing full-length cDNA and use thereof  
JOURNAL Patent: JP 2002191363-A 5189 09-JUL-2002;  
COMMENT  
OS Homo sapiens (human)  
PN JP 2002191363-A/5189  
PD 09-JUL-2002  
PF 28-JUL-2000 JP 2002280990  
PI TOSHIO OTA,TAKAO ISOGAI,TETSUO NISHIKAWA,KOJI HAYASHI,KAORI  
PI SAITO,  
PI JUNICHI YAMAMOTO,SHIZUKO ISHII,TOMOYASU SUGIYAMA,AI WAKAMATSU,  
PI KEIICHI NAGAI,TETSUJI OTSUKI

```

PC
C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC
10, C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key
Location/Qualifiers
FT source 1. 680
/organism='Homo sapiens (human)'
FEATURES
source
1. 680
/organism='Homo sapiens'
/mol_type='genomic DNA'
/db_xref='taxon:9606'
ORIGIN
Query Match 80.0%; Score 33.6; DB 2; Length 680;
Best Local Similarity 87.8%; Pred. No. 0.0007;
Matches 36; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 GCCACGCTTGCCGAGCCCTTACAAAGACTATGAAGTAA 41
|||||
Db 633 GCCACGCTTGCCCAACCTTACAAAGACTATGAAGTAA 673
|||||

RESULT 23
AX870284
LOCUS
AX870284
DEFINITION
Sequence 5189 from Patent EP1074617.
ACCESSION
AX870284
VERSION
AX870284.1 GI:40025147
KEYWORDS
Homo sapiens (human)
SOURCE
Homo sapiens
ORGANISM
Homo sapiens
REFERENCE
1 Ota, T., Iwagai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,
Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.
Primers for synthesizing full-length cDNA and their use
Patent: EP 1074617-A 5189 07-FEB-2001;
Research Association for Biotechnology (JP)
Location/Qualifiers
FEATURES
source
1. 680
/organism='Homo sapiens'
/mol_type='unassigned DNA'
/db_xref='taxon:9606'
ORIGIN
Query Match 80.0%; Score 33.6; DB 2; Length 680;
Best Local Similarity 87.8%; Pred. No. 0.0007;
Matches 36; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 GCCACGCTTGCCGAGCCCTTACAAAGACTATGAAGTAA 41
|||||
Db 633 GCCACGCTTGCCCAACCTTACAAAGACTATGAAGTAA 673
|||||

RESULT 24
AC132640
LOCUS
AC132640
DEFINITION
Rattus norvegicus clone CH230-339B14, WORKING DRAFT SEQUENCE.
ACCESSION
AC132640
VERSION
HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS
Rattus norvegicus (Norway rat)
SOURCE
Rattus norvegicus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Rattus
REFERENCE
1 (bases 1 to 194709)
Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,

```

```

Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blait, J., Blankenburg, K., Blych, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Caldwell, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
Chacko, J., Chavez, D., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Devila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hlyak, S. L., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Huiy, K., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorensuhewa, L., Loulsegged, H., Lozado, R. J., Lu, X., Ma, J.,
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Naif, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwaokemele, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkoch, C.,
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.,
Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,
Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,
Weinstock, G. and Gibbs, R. A.
Direct Submission
Unpublished
2 (bases 1 to 194709)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (02-SEP-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 194709)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 15, 2002 this sequence version replaced gi:23908481.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature

```

```

table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: KBQF
Center clone name: CH230-339B14
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 189047 bases at least Q40
Consensus quality: 189447 bases at least Q30
Consensus quality: 190412 bases at least Q20
Estimated insert size: 195068; sum-of-contigs estimation
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 194709: contig of 194709 bp in length.
* Location/Qualifiers
* 1..194709
* /organism="Rattus norvegicus"
* /mol_type="genomic DNA"
* /db_xref="taxon:10116"
* /clone="CH230-339B14"
* 1..3065
* /note="wgs_end_extension
clone_end:T7"
7050..7929
/note="clone_boundary
clone_end:T7"
site:
end_sequence:BZ164292"
complement(191365..192203)
/note="clone_boundary
clone_end:Sp6
site:
end_sequence:BZ164293"

FEATURES
source
misc_feature
misc_feature
misc_feature

ORIGIN
Query Match 61.9%; Score 26; DB 12; Length 194709;
Best Local Similarity 76.2%; Pred. No. 16;
Matches 32; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 1 GCCACGGTGGCCAGCGCTTACAAAGACTATGAGTAAG 42
Db 169254 GCCAACTGTGGCCAGTCCAGCAAGGACAGGAAGTAAG 169295

RESULT 25
AC108634
LOCUS
DEFINITION Rattus norvegicus clone CH230-249H12, *** SEQUENCING IN PROGRESS
***, 5 unordered pieces.
AC108634
ACCESSION AC108634.4 GI:231195538
VERSION HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Rattus.
1 (bases 1 to 232234)

```

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information  
Center project name: KBQF  
Center clone name: CH230-339B14  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 189047 bases at least Q40  
Consensus quality: 189447 bases at least Q30  
Consensus quality: 190412 bases at least Q20  
Estimated insert size: 195068; sum-of-contigs estimation  
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 1 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submittor.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\* 1 194709: contig of 194709 bp in length.  
\* Location/Qualifiers  
\* 1..194709  
\* /organism="Rattus norvegicus"  
\* /mol\_type="genomic DNA"  
\* /db\_xref="taxon:10116"  
\* /clone="CH230-339B14"  
\* 1..3065  
\* /note="wgs\_end\_extension  
clone\_end:T7"  
7050..7929  
/note="clone\_boundary  
clone\_end:T7"  
site:  
end\_sequence:BZ164292"  
complement(191365..192203)  
/note="clone\_boundary  
clone\_end:Sp6  
site:  
end\_sequence:BZ164293"

FEATURES  
source  
misc\_feature  
misc\_feature  
misc\_feature

ORIGIN  
Query Match 61.9%; Score 26; DB 12; Length 194709;  
Best Local Similarity 76.2%; Pred. No. 16;  
Matches 32; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 1 GCCACGGTGGCCAGCGCTTACAAAGACTATGAGTAAG 42  
Db 169254 GCCAACTGTGGCCAGTCCAGCAAGGACAGGAAGTAAG 169295

RESULT 25  
AC108634  
LOCUS  
DEFINITION Rattus norvegicus clone CH230-249H12, \*\*\* SEQUENCING IN PROGRESS  
\*\*\*, 5 unordered pieces.  
AC108634  
ACCESSION AC108634.4 GI:231195538  
VERSION HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.  
KEYWORDS Rattus norvegicus (Norway rat)  
SOURCE Rattus norvegicus  
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Rattus.  
1 (bases 1 to 232234)

AUTHORS  
Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,  
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,  
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,  
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,  
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,  
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,  
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,  
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,  
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,  
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,  
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,  
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
Fregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,  
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,  
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
Hernandez, R., Hines, S., Hladun, S. B., Hodgson, A., Hogue, M.,  
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,  
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,  
Karpachy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,  
Kovis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
Lorenshew, L., Louissegh, H., Lozano, R. J., Lu, X., Ma, J.,  
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,  
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,  
Mawhinney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,  
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,  
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,  
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K.,  
Nwaokemele, O., Okwuonu, G., Olarnpungsoo, A., Pal, S., Parks, K.,  
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,  
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,  
Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,  
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
Rives, C., Rodkey, T., Rojas, A., Rose, R., Ruiz, S. B.,  
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,  
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D.,  
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,  
Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,  
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,  
Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,  
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,  
Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K.,  
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
Yu, F., Zhang, J., Zhou, X., Zhou, X., Zhao, S., Dunn, D., von  
Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,  
Weinstock, G. and Gibbs, R. A.  
Direct Submission  
Unpublished  
2 (bases 1 to 232234)  
Worley, K. C.  
Direct Submission  
Submitted (31-JAN-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 232234)  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (19-SEP-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
The sequence in this assembly is a combination of BAC based reads  
and whole genome shotgun sequencing reads assembled using Atlas  
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). As a result, the  
sequence may extend beyond the ends of the clone and there may be  
contigs that consist entirely of whole genome shotgun sequence  
reads. Both end sequences and whole genome shotgun sequence only  
contigs will be indicated in the feature table.  
----- Genome Center  
Center: Baylor College of Medicine



```

Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GPQV
Center clone name: CH230-249H12
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 213554 bases at least Q40
Consensus quality: 216011 bases at least Q30
Consensus quality: 217771 bases at least Q20
Estimated insert size: 236418; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 224930: contig of 224930 bp in length
* 224931 225030: gap of unknown length
* 225031 228442: contig of 3412 bp in length
* 228443 228542: gap of unknown length
* 228543 229623: contig of 1081 bp in length
* 229624 229723: gap of unknown length
* 229724 231059: contig of 1336 bp in length
* 231060 231159: gap of unknown length
* 231160 232234: contig of 1075 bp in length.
*
* Location/Qualifiers
* 1. 232234
* /organism="Rattus norvegicus"
* /mol_type="genomic DNA"
* /db_xref="taxon:10116"
* /clone="CH230-249H12"
*
* misc_feature
* 1. 1150
* /note="wgs_contig"
*
* gap
* 224931..225030
* /estimated_length=unknown
*
* misc_feature
* 225031..226918
* /note="wgs_contig"
*
* gap
* 228443..228542
* /estimated_length=unknown
*
* gap
* 229624..229723
* /estimated_length=unknown
*
* gap
* 231060..231159
* /estimated_length=unknown
*
* ORIGIN
*
* Query Match 61.9%; Score 26; DB 12; Length 232234;
* Best Local Similarity 76.2%; Pred. No. 17;
* Matches 32; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
*
* QY 1 GCCACGCTTGCCGAGCCCTTCAAGACTATGAAGTAAAG 42
* |||||
* DB 130851 GCCAACTGTGGCCAGTCCAGACAGCAAGTAAAG 130892
* |||||
*
* RESULT 26
* AC107096/c
* AC107096 LOCUS
* DEFINITION Rattus norvegicus clone CH230-229H11, *** SEQUENCING IN PROGRESS
* ** 3 unordered pieces.
* AC107096
* AC107096.5 GI:30580079
* HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
* SOURCE Rattus norvegicus (Norway rat)
* ORGANISM Rattus norvegicus
* Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

REFERENCE  
AUTHORS

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Murioidea; Muridae; Murinae; Rattus.

1 (bases 1 to 240216)

Muzny, D., Marie, Metzker, M. Lee., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, S., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escoto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuewa, L., Loulseged, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhiney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K., Nwaokeme, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Plannkoch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, P., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajz, D., Speed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

Direct Submission  
Unpublished

2 (bases 1 to 240216)

Worley, K. C.

Direct Submission

Submitted (16-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 240216)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On May 13, 2003 this sequence version replaced gi:23116012.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated

TITLE  
JOURNAL

REFERENCE  
AUTHORS

TITLE  
JOURNALREFERENCE  
AUTHORSTITLE  
JOURNAL

## COMMENT

by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

----- Project Information  
Center project name: GKDH

Center clone name: CH230-229H11

----- Summary Statistics

Assembly program: Atlas 3.0;

Consensus quality: 228704 bases at least Q40

Consensus quality: 230891 bases at least Q30

Consensus quality: 232798 bases at least Q20

Estimated insert size: 242082; sum-of-contigs estimation

Quality coverage: 6X in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 3 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 14751: contig of 14751 bp in length

\* 14752 14851: gap of unknown length

\* 14852 173383: contig of 158532 bp in length

\* 173384 173483: gap of unknown length

\* 173484 240216: contig of 66733 bp in length.

## FEATURES

## source

1. 240216

/organism="Rattus norvegicus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10116"

/clone="CH230-229H11"

1. 1213

/note="wgs\_end\_extension"

clone\_end:Sp6"

complement(1109..62776)

/note="clone boundary"

clone\_end:Sp6"

site:ECORI

end\_sequence:B2111003"

2320..2652

/note="clone boundary"

clone\_end:T7"

site:ECORI

end\_sequence:B2111001"

4147..5522

/note="wgs\_end\_extension"

clone\_end:T7"

14752..14851

/estimated\_length:unknown

103697..105370

/note="wgs\_end\_extension"

clone\_end:T7"

173384..173483

/estimated\_length:unknown

## ORIGIN

Query Match 61.9%; Score 26; DB 12; Length 240216;

Best Local Similarity 76.2%; Pred. No. 17;

Matches 32; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 GCCACGGTTGGCCGCGCCCTTACAAAGACTATGAGTAAG 42

Db 121757 GCCTACTGTTGGCCCGCTCCAGACAGGACAGGAAGTAAG 121716

RESULT 27  
AC105586

DEFINITION  
Rattus norvegicus 3 BAC CH230-239D4 (Children's Hospital Oakland Research Institute) complete sequence.

AC105586

VERSION  
AC105586.5 GI:31376405

KEYWORDS  
HTG.

SOURCE  
Rattus norvegicus (Norway rat)

ORGANISM  
Rattus norvegicus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muroidea; Muridae; Murinae; Rattus.

1 (bases 1 to 235468)

Murphy, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,

Albrooks, S.L., Amaratunga, H.C., Are, J.R., Ayale, M., Banks, T.,

Barbora, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D.,

Bouck, J., Bowie, S., Brieve, M., Brown, E., Brown, M., Bryant, N.P.,

Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,

Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,

Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,

Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,

Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,

Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,

Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,

Earnhart, C., Edgar, B., Edwards, C.C., Elhaj, C., Escotto, M.,

Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,

Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,

Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,

Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., He, X.,

Hernandez, J., Hernandez, O., Hodgson, A., Hughes, M., Holloway, C.,

Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,

Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,

Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,

Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,

Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,

Louise, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,

Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,

Martinez, B., Massey, E., Mawhinney, E., McLeod, M.P., Meador, M.,

Mei, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,

Moore, S., Morgan, M., Moorish, T., Morris, S., Moser, M., Neal, D.,

Nelson, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,

Nickerson, E., Nwokenkwo, S., Ogur, M., Okwuonu, G., Oragunye, N.,

Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L.,

Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M.,

Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S.,

Scott, G., Shen, H., Shoohtari, N., Sisson, I., Sodergren, E.,

Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A.,

Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C.,

Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L.,

Vera, V., Villalon, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S.,

Warren, R., Washington, C., Watlington, S., Williams, G.,

Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y.,

Wu, Y.F., Zhou, J., Zorrilla, S., Zorrilla, S.L., Weinstein, G. and

Gibbs, R.

Direct Submission

TITLE

JOURNAL

REFERENCE

2 (bases 1 to 235468)

Worley, K.C.

Direct Submission

TITLE

JOURNAL

REFERENCE

3 (bases 1 to 235468)

Worley, K.C.

Direct Submission

TITLE

JOURNAL

REFERENCE

3 (bases 1 to 235468)

Worley, K.C.

Direct Submission

TITLE

JOURNAL

REFERENCE



DEFINITION	Rattus norvegicus clone CH230-7M5, WORKING DRAFT SEQUENCE.	COMMENT	
ACCESSION	AC126155		
VERSION	AC126155.4 GI:30579514		
KEYWORDS	HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.		
SOURCE	Rattus norvegicus (Norway rat)		
ORGANISM	Rattus norvegicus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Murioidea; Muridae; Murinae; Rattus.		
AUTHORS	1 (bases 1 to 253631) Muzny, D. Marie., Metzker, M. Lee., Abramzon, S., Adams, C., Alder, J., Allen, C., Allen, H., Alebrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Ayogaji, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Frazer, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, I., Garza, M., Gebregorgis, E., Geier, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogue, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Loresnuhewa, L., Louised, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangun, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Miosavijevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaekelemeh, O., Okwunou, G., Olarpunsaagon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Popper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L. L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, R., Thomas, S., Tingey, A., Trejos, Z., Umani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlarczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.		
	Unpublished		
JOURNAL	2 (bases 1 to 253631)		
REFERENCE	Worley, K. C.		
AUTHORS	Submitted (04-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA		
JOURNAL	3 (bases 1 to 253631)		
REFERENCE	Rat Genome Sequencing Consortium.		
AUTHORS	Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA		
JOURNAL			
On May 13, 2003 this sequence version replaced gi:22856732.			
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.			
----- Genome Center			
Center: Baylor College of Medicine			
Center code: BCM			
Web site: http://www.hgsc.bcm.tmc.edu/			
Contact: hgsc-help@bcm.tmc.edu			
----- Project Information			
Center project name: GKAR			
Center clone name: CH230-7M5			
----- Summary Statistics			
Assembly program: Atlas 3.0;			
Consensus quality: 245168 bases at least Q40			
Consensus quality: 246824 bases at least Q30			
Consensus quality: 247704 bases at least Q20			
Estimated insert size: 254802; sum-of-contigs estimation			
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation			
-----			
* NOTE: Estimated insert size may differ from sequence length			
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)			
* NOTE: This sequence may represent more than one clone.			
* NOTE: This is a 'working draft' sequence. It currently			
* consists of 1 contigs. Gaps between the contigs			
* are represented as runs of N. The order of the pieces			
* is believed to be correct as given, however the sizes			
* of the gaps between them are based on estimates that have			
* provided by the submittor.			
* This sequence will be replaced			
* by the finished sequence as soon as it is available and			
* the accession number will be preserved.			
* 1 253631: contig of 253631 bp in length.			
Location/Qualifiers			
1..253631			
/organism="Rattus norvegicus"			
/mol_type="genomic DNA"			
/db_xref="taxon:10116"			
/clone="CH230-7M5"			
FEATURES			
source			
ORIGIN			
Query Match	58.6%;	Score 24.6;	DB 12; Length 253631;
Best Local Similarity	76.9%;	Pred. No. 77;	
Matches	30; Conservative	0; Mismatches	9; Indels 0; Gaps 0;
QY	3 CCACGCTTGGCGGCCCTTACAAAGACTATGAGCTATAA 41		
DB	37759 CCATGGTGGCAAGCCCTGACAAAGACTAGGAATGAA 37797		
RESULT 29			
AC116169	AC116169	189833 bp	DNA linear HTG 27-MAR-2003
DEFINITION	Homo sapiens clone RP11-55C13, 2 ordered pieces.		
ACCESSION	AC116169		
VERSION	AC116169.5 GI:23129007		
KEYWORDS	HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_CANCELLED.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (bases 1 to 189833)		
AUTHORS	Birren, B., Nusbaum, C. and Lander, E.		



Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleciyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission  
Unpublished  
2 (bases 1 to 229303)  
Worley, K.C.

Direct Submission  
Submitted (07-FEB-2005) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 229303)  
Cow Genome Sequencing Consortium.

Direct Submission  
Submitted (01-JUL-2005) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On June 29, 2005 this sequence version replaced gi:58743475.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: FDOO  
Center clone name: CH240-62M20  
----- Summary Statistics  
Assembly program: Atlas 3.0;  
Consensus quality: 213349 bases at least Q40  
Consensus quality: 215963 bases at least Q30  
Consensus quality: 218476 bases at least Q20  
Estimated insert size: 220573; sum-of-contigs estimation  
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently consists of 27 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.  
\* 1 9337: contig of 9337 bp in length  
\* 9338 10432: gap of 1095 bp  
\* 10433 13724: contig of 3292 bp in length  
\* 13725 13774: gap of 50 bp  
\* 13775 16987: contig of 3213 bp in length  
\* 16988 17087: gap of unknown length  
\* 17088 20724: contig of 3637 bp in length  
\* 20725 20774: gap of 50 bp

20775 26078: contig of 5304 bp in length  
26079 26128: gap of 50 bp  
26129 28649: contig of 2521 bp in length  
28650 29631: gap of 981 bp  
29631 34670: contig of 5040 bp in length  
34671 34720: gap of 50 bp  
34721 40011: contig of 5291 bp in length  
40012 40111: gap of unknown length  
40112 89022: contig of 48861 bp in length  
89023 89022: gap of 50 bp  
89023 111089: contig of 22067 bp in length  
111090 111139: gap of 50 bp  
111140 131191: contig of 20052 bp in length  
131192 132275: gap of 1084 bp  
132276 133460: contig of 1185 bp in length  
133461 133560: gap of unknown length  
133561 137385: contig of 3825 bp in length  
137386 137435: gap of 50 bp  
137436 144746: contig of 7311 bp in length  
144747 144846: gap of unknown length  
144847 145973: contig of 1127 bp in length  
145974 146023: gap of 50 bp  
146024 148479: contig of 2456 bp in length  
148480 148579: gap of unknown length  
148580 148669: contig of 1290 bp in length  
149870 150090: gap of 221 bp  
150091 160007: contig of 9917 bp in length  
160008 160057: gap of 50 bp  
160058 163938: contig of 3881 bp in length  
163939 163988: gap of 50 bp  
163989 177724: contig of 13736 bp in length  
177725 177774: gap of 50 bp  
177775 179811: contig of 2037 bp in length  
179812 180606: gap of 795 bp  
180607 194799: contig of 14193 bp in length  
194800 194849: gap of 50 bp  
194850 209924: contig of 15075 bp in length  
209925 209974: gap of 50 bp  
209975 215978: contig of 6004 bp in length  
215979 216028: gap of 50 bp  
216029 224039: contig of 8011 bp in length  
224040 225433: gap of 1394 bp  
225434 228175: contig of 2742 bp in length  
228176 228275: gap of unknown length  
228276 229303: contig of 1028 bp in length.

FEATURES  
source  
1. 229303  
/organism="Bos taurus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9913"  
/clone="CH240-62M20"  
9338. .10432  
/estimated\_length=1095  
13725. .13774  
/estimated\_length=50  
16988. .17087  
/estimated\_length=unknown  
20725. .20774  
/estimated\_length=50  
26079. .26128  
/estimated\_length=50  
28650. .29630  
/estimated\_length=981  
34671. .34720  
/estimated\_length=50  
40012. .40111  
/estimated\_length=unknown  
89023. .89022  
/estimated\_length=50  
111090. .111139  
/estimated\_length=50  
131192. .132275  
/estimated\_length=1084







```
* the accession number will be preserved.
```

```
*      1   169003: contig of 169003 bp in length.
```

```
FEATURES
```

```
source              Location/Qualifiers
```

```
1..169003
```

```
/organism="Danio rerio"
```

```
/mol_type="genomic DNA"
```

```
/db_xref="taxon:7955"
```

```
/chromosome="13"
```

```
/clone="CH211-37A16"
```

```
/clone_lib="CHORI-211"
```

```
misc_feature        1..169003
```

```
/note="assembly_fragment:00015"
```

```
clone_end:SP6
```

```
vector_side:left"
```

```
ORIGIN
```

```
Query Match          54.8%; Score 23; DB 12; Length 169003;
```

```
Best Local Similarity 74.4%; Pred. No. 3.7e+02;
```

```
Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
```

```
QY      3 CCACGGCTTCGCGAGCCCTTACAAGACTATGAAGTAAA 41
```

```
Dd      137081 CTCGCTTTTGCTTACCACATTACAAGCTATAGAATATA 137043
```

```
RESULT 34
```

```
AC159967
```

```
LOCUS       AC159967               185146 bp    DNA     linear   ROD 10-AUG-2005
```

```
DEFINITION  Mus musculus chromosome 1, clone RP23-302P9, complete sequence.
```

```
ACCESSION   AC159967
```

```
VERSION     AC159967.2 GI:72096116
```

```
KEYWORDS    HTG.
```

```
SOURCE      Mus musculus (house mouse)
```

```
ORGANISM    Mus musculus
```

```
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Muridae; Murinae; Mus.
```

```
REFERENCE   1 (bases 1 to 185146)
```

```
AUTHORS     Birren,B., Nusbaum,C. and Lander,E.
```

```
TITLE       Mus musculus chromosome 1, clone RP23-302P9
```

```
JOURNAL     Unpublished
```

```
REFERENCE   2 (bases 1 to 185146)
```

```
AUTHORS     Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhalter,B., Camarata,J., Chang,J., Chopel,Y., Collamore,A., Cook,A., Cooke,P., Corum,B., DeArrellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-pierre,N., Hafez,N., Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R., MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M., McDrim,J., Meneus,L., Mihova,T., Milenga,V., Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M., Talanas,J., Testaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Venkataman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembeke,L., Zimmer,A. and Zody,M.
```

```
Direct Submission
```

```
Submitted (14-APR-2005) Broad Institute of MIT and Harvard, 320 Charles Street, Cambridge, MA 02141, USA
```

```
3 (bases 1 to 185146)
```

```
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhalter,B., Camarata,J., Chang,J., Chopel,Y., Collamore,A., Cook,A., Cooke,P., Corum,B., DeArrellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R., MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M., McDrim,J., Meneus,L., Mihova,T., Milenga,V., Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M., Talanas,J., Testaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Venkataman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembeke,L., Zimmer,A. and Zody,M.
```

```
TITLE       Direct Submission
```

```
JOURNAL     Submitted (14-APR-2005) Broad Institute of MIT and Harvard, 320 Charles Street, Cambridge, MA 02141, USA
```

```
REFERENCE   3 (bases 1 to 185146)
```

```
AUTHORS     Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhalter,B., Camarata,J., Chang,J., Chopel,Y., Collamore,A., Cook,A., Cooke,P., Corum,B., DeArrellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R., MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M., McDrim,J., Meneus,L., Mihova,T., Milenga,V., Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M., Talanas,J., Testaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Venkataman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembeke,L., Zimmer,A. and Zody,M.
```

Galgan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Liu, X., Lui, A., Mabbitt, R., MacLean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrum, J., Menelus, L., Minova, T., Mieng, V., Murphy, T., Naylor, J., Nguyen, C., Nguyen, T., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zemбек, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (10-AUG-2005) Broad Institute of MIT and Harvard, 320 Charles Street, Cambridge, MA 02141, USA

On Aug 10, 2005 this sequence version replaced 91:62543423.

All repeats were identified using RepeatMasker:  
Smt, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center -----

Center: Broad Institute of MIT and Harvard  
Center code: W1BR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@broad.mit.edu](mailto:sequence_submissions@broad.mit.edu)

----- Project Information -----

Center project name: J5628  
Center clone name: 302\_P\_9

-----

repeat_region	complement (8323. .3649)
repeat_region	/rpt family="L1_Mus3"
repeat_region	10409. .10519
repeat_region	/rpt family="MTD"
repeat_region	10518. .10617
repeat_region	/rpt family="B1_Mur2"
repeat_region	11110. .11504
repeat_region	/rpt family="MTB_Mm"
repeat_region	11625. .11769
repeat_region	/rpt family="RSINE1"
repeat_region	11911. .12075
repeat_region	/rpt family="Lx3B"
repeat_region	complement (12727. .12866)
repeat_region	/rpt family="B1_Mur2"
repeat_region	complement (13372. .13559)
repeat_region	/rpt family="RMER15"
repeat_region	14231. .14265
repeat_region	/rpt family="(A)"
repeat_region	complement (14284. .14404)
repeat_region	/rpt family="B1F2"
repeat_region	complement (14806. .14900)
repeat_region	/rpt family="PB1D7"
repeat_region	complement (15044. .15250)
repeat_region	/rpt family="B3"
repeat_region	15689. .16181
repeat_region	/rpt family="MT2_Mm"
repeat_region	16182. .16890
repeat_region	/rpt family="MERVL_Mm"
repeat_region	16886. .21308
repeat_region	/rpt family="MERVL_Mm"
repeat_region	21316. .21402
repeat_region	/rpt family="ORR1A3-int"
repeat_region	21403. .21440
repeat_region	/rpt family="MERVL_Mm"
repeat_region	21441. .21933
repeat_region	/rpt family="MT2_Mm"
repeat_region	23018. .23181
repeat_region	/rpt family="B4A"
repeat_region	23247. .23413
repeat_region	/rpt family="MTEa"
repeat_region	23712. .23848
repeat_region	/rpt family="B1_Mur3"
repeat_region	23916. .24524
repeat_region	/rpt family="RMER1C"
repeat_region	complement (24525. .24836)
repeat_region	/rpt family="ORR1A4"
repeat_region	24837. .24866
repeat_region	/rpt family="RMER1C"
repeat_region	complement (24899. .25068)
repeat_region	/rpt family="Lx7"
repeat_region	25075. .25108
repeat_region	/rpt family="(CA)"
repeat_region	complement (25115. .25612)
repeat_region	/rpt family="Lx7"
repeat_region	25642. .26722
repeat_region	/rpt family="RSINE1"
repeat_region	26808. .26828
repeat_region	/rpt family="GC_rich"
repeat_region	26812. .26832
repeat_region	/rpt family="GC_rich"
repeat_region	complement (27289. .27365)
repeat_region	/rpt family="MLT1L"
repeat_region	27479. .27563
repeat_region	/rpt family="(GA)"
repeat_region	27618. .27771
repeat_region	/rpt family="RSINE1"
repeat_region	27842. .28036
repeat_region	/rpt family="ID_B1"
repeat_region	28054. .28108
repeat_region	/rpt family="ID4"
repeat_region	complement (28312. .28635)
repeat_region	/rpt family="MTE2b"
repeat_region	complement (28626. .28706)

```

repeat_region      /rpt_family="MTE2b"
repeat_region      /rpt_family="B3"
repeat_region      /rpt_family="B4A"
repeat_region      /rpt_family="(CA)n"
repeat_region      30898..30933

Query Match      54.8%; Score 23; DB 6; Length 185146;
Best Local Similarity 83.9%; Pred. No. 3.8e+02;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 CCACGCTTGGCGAGCGCTTACAAAGACTAT 33
Db 169261 CCACCTTGGCGAGCGCTTGCACAGGTAT 169291

RESULT 35
AC102564
LOCUS      230474 bp DNA linear ROD 21-JAN-2005
DEFINITION Mus musculus chromosome 1, clone RP23-210C12, complete sequence.
ACCESSION AC102564
VERSION AC102564.7 GI:58000610
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 230474)
AUTHORS Birren,B., Nusbaum,C. and Lander,E.
TITLE Mus musculus chromosome 1, clone RP23-210C12
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 230474)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Chospel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
Lamazares,R., Landers,T., Lenoczky,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marguis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 230474)
AUTHORS Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,
Bloom,T., Boguslavsky,L., Boukhgalter,B., Camarata,J., Chang,J.,
Chospel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,
DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,
Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D.,
Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,
Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,
MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,

```

```

Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,
Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,
Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (10-DEC-2004) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 4 (bases 1 to 230474)
AUTHORS Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,
Bloom,T., Boguslavsky,L., Boukhgalter,B., Camarata,J., Chang,J.,
Chospel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,
DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,
Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D.,
Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,
Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,
MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,
Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,
Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (21-JAN-2005) Broad Institute of MIT and Harvard, 320
Charles Street, Cambridge, MA 02141, USA
COMMENT On Jan 21, 2005 this sequence version replaced gi:55831552.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Broad Institute of MIT and Harvard
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@broad.mit.edu
----- Project Information
Center project name: L19015
Center clone name: 210_C_12
-----
FEATURES
source
Location/Qualifiers
1. 230474
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="1"
/map="1"
/clone="RP23-210C12"
/clone_lib="RPCI-23 Female Mouse BAC"
complement(4..80)
/rpt_family="PB1D9"
complement(101..307)
/rpt_family="B3"
484..519
/rpt_family="(CATA)n"
657..888
/rpt_family="B3A"
960..1002
/rpt_family="(TCCC)n"
complement(1005..1147)
/rpt_family="B1_MM"
1306..1387
/rpt_family="ID_B1"
repeat_region
repeat_region
repeat_region
repeat_region
repeat_region
repeat_region
repeat_region

```

```
repeat_region 1468..1491
/rpt_family="AT-rich"
repeat_region complement(2746..2827)
/rpt_family="MIR"
repeat_region 3070..3313
/rpt_family="L2"
repeat_region complement(3536..3591)
/rpt_family="MIR"
repeat_region 3985..4358
/rpt_family="ORR1D"
repeat_region 5146..5265
/rpt_family="RSINE1"
repeat_region 5266..5297
/rpt_family=" (CTCA)n"
repeat_region 5308..5353
/rpt_family=" (CAGA)n"
repeat_region 5537..5727
/rpt_family="B2_Mm1"
repeat_region 5731..5877
/rpt_family="B1_MM"
repeat_region 5880..5949
/rpt_family="GA-rich"
repeat_region 6044..6181
/rpt_family="B1_MM"
repeat_region 6182..6218
/rpt_family=" (CAAAA)n"
repeat_region 7488..7586
/rpt_family="L2"
repeat_region complement(7628..7722)
/rpt_family="MIR"
repeat_region 8385..8541
/rpt_family="B3A"
repeat_region 8548..8590
/rpt_family=" (TCTG)n"
repeat_region 8604..8673
/rpt_family="B1F"
repeat_region 8674..8731
/rpt_family=" (GA)n"
repeat_region 8732..8762
/rpt_family="ID_B1"
repeat_region 9094..9219
/rpt_family="MER34C"
repeat_region complement(9374..9492)
/rpt_family="PB1"
repeat_region complement(10096..10217)
/rpt_family="Lx6"
repeat_region 10498..10747
/rpt_family="B4A"
repeat_region 10748..10870
/rpt_family="RSINE1"
repeat_region 10897..10923
/rpt_family=" (CAAA)n"
repeat_region 10979..11028
/rpt_family=" (CATA)n"
repeat_region 11079..11154
/rpt_family="B1F"
repeat_region 11155..11174
/rpt_family="ID_B1"
repeat_region 11241..11311
/rpt_family=" (TTTC)n"
repeat_region complement(11313..11471)
/rpt_family="B1_MM"
repeat_region 12420..12442
/rpt_family=" (TTTA)n"
repeat_region 12442..12509
/rpt_family=" (TG)n"
repeat_region complement(12519..12803)
/rpt_family="B4A"
repeat_region 12863..12911
/rpt_family=" (TA)n"
repeat_region 13046..13099
/rpt_family="B1_MM"
repeat_region 13106..13263
```

---

```
Query Match 54.8%; Score 23; DB 6; Length 230474;
Best Local Similarity 83.9%; Pred. No. 4.1e+02;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 CCACGCTTGGCGGAGCCCTTACAAAGACTAT 33
DB 94133 CCACCTTGGCGGAGCCCTTGCCAAAGTTAT 94163

RESULT 36
LOCUS AR630602 1500 bp DNA linear PAT 14-FEB-2005
DEFINITION Sequence 3 from patent US 6841374.
ACCESSION AR630602
VERSION AR630602.1 GI:59765853
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1500)
AUTHORS Ishikawa,T., Yoshi,H., Onishi,T., Imagawa,T. and Ishibaashi,M.
TITLE Enhanced immunogen for inactivated vaccine for infection with
Japanese encephalitis viruses and process for producing the same
Patent: US 6841374-A 3 11-JAN-2005;
Research Foundation for Microbial Diseases of Osaka University;
Osaka;
JPX;

FEATURES             Location/Qualifiers
     source           1..1500
                     /organism="unknown"
                     /mol_type="genomic DNA"

ORIGIN
Query Match 54.3%; Score 22.8; DB 2; Length 1500;
Best Local Similarity 79.4%; Pred. No. 88;
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 CACGCTTGGCGGAGCCCTTACAAAGACTATGAAG 37
DB 1203 CACGCTGGGCAAGCCCTTTCAACGACTTTGAAG 1236

RESULT 37
AF112297 LOCUS AF112297 1500 bp RNA linear VRL 12-APR-1999
DEFINITION Japanese encephalitis virus (STRAIN NAKAYAMA) envelope protein
ACCESSION AF112297
VERSION AF112297.1 GI:4580915
KEYWORDS
SOURCE Japanese encephalitis virus strain Nakayama
ORGANISM Japanese encephalitis virus strain Nakayama
REFERENCE 1 (bases 1 to 1500)
AUTHORS Chin,C., Sung,S.-C. and Chang,G.-J.
TITLE Flavivirus; Japanese encephalitis virus group.
TITLE Genetic variation of envelope protein gene of Japanese encephalitis
virus
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1500)
AUTHORS Chin,C., Sung,S.-C. and Chang,G.-J.
TITLE Direct Submission
JOURNAL Submitted (08-DEC-1998) Division of Vector-Borne Infectious
Diseases, Centers for Disease Control and Prevention, P.O. Box
2087, Fort Collins, CO 80522, USA

FEATURES             Location/Qualifiers
     source           1..1500
                     /organism="Japanese encephalitis virus strain Nakayama"
                     /mol_type="genomic RNA"
                     /strain="Nakayama"
                     /db_xref="taxon:11076"
```

```

/notes="from Walter Reed Army Institute of Research"
<1..>1500
/codon_start=1
/product="envelope protein"
/protein_id="AAD24538.1"
/db_xref="GI:4580916"
/translation="FNCILGMGNRDFIEGASGATWVDLVLEGDSCLTLMANDKPTLDVR
MINIEAVQLAEVRSYCYHASVTDISTVARCPTTGAHNKRRADSSVYCKQGFTRDGMG
NCGGLFGKSIDTCAKFSCTKAIGRTIQPENIKYEVGIFVHGTTTSENHGNYSQVG
ASQAAKFTVTNAPSITLKLGDYGEVTLDCPRSLNTEAFYVMTVSGSKFLVHREWF
HDALPFTPPSTAWNRRELLMESEAHATKQSVVALGQEGGLHQALAGAIIVVEYSN
SVKLTSGHLKCRKMDKALKGTTVMCTEKFSPAKNPADTGHGTVIELSGSDGP
CKIPIVSVASLNDMTPVGLTVNPFVATSSANSKVLVEMEPFGDSYIVVGRGDKQI
NHHMKAGSTLTKAFSTTLKGAORLAALGDTAWDFGSGGVFNSIGKAVHQVFGGAFR
TLFGGMSWITQGLMGALLMGMGNARDRSIALAFLATGGLVFLATNVHA"

ORIGIN
Query Match      54.3%; Score 22.8; DB 10; Length 1500;
Best Local Similarity 79.4%; Pred. No. 88;
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 CACGCTTGCCGAGCCCTTACAAAGACTATGAAG 37
    ||||| ||| ||||| ||||| ||||| |||||
Db 1203 CACGCTGGGCAAGCCTTTTCAACGACTTTGAAG 1236

RESULT 39
AY243826 Japanese encephalitis virus isolate Tla polyprotein gene, partial
LOCUS AY243826 1500 bp RNA linear VRL 27-APR-2003
DEFINITION cds.
ACCESSION AY243826
VERSION AY243826.1 GI:30143712
KEYWORDS
SOURCE
ORGANISM Japanese encephalitis virus
Japanese encephalitis virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Flavivirus; Japanese encephalitis virus group.
REFERENCE 1 (bases 1 to 1500)
AUTHORS Liang, G., Li, X. and Fu, S.
TITLE Direct Submission
JOURNAL Submitted (25-FEB-2003) State Key Lab. of Molecular Virology &
Genetic Engineering, Inst. of Virology, 100 Yingxin St., Xuan Wu
Qu, Beijing 100052, China
FEATURES
source
location/Qualifiers
1..1500
/organism="Japanese encephalitis virus"
/mol_type="genomic RNA"
/isolate="Tla"
/isolation_source="patient brain sample"
/db_xref="taxon:11072"
/country="China"
<1..>1500
/notes="envelope protein region"
/codon_start=1
/product="polyprotein"
/protein_id="AAP14877.1"
/db_xref="GI:30143713"

CDS
translation="FNCILGMGNRDFIEGASGATWVDLVLEGDSCLTIMANDKPTLDVR
MINIEAVQLAEVRSYCYHASVTDISTVARCPTTGAHNKRRADSSVYCKQGFTRDGMG
NCGGLFGKSIDTCAKFSCTKAIGRTIQPENIKYEVGIFVHGTTTSENHGNYSQVG
ASQAAKFTVTNAPSITLKLGDYGEVTLDCPRSLNTEAFYVMTVSGSKFLVHREWF
HDALPFTPPSTAWNRRELLMESEAHATKQSVVALGQEGGLHQALAGAIIVVEYSN
SVKLTSGHLKCRKMDKALKGTTVMCTEKFSPAKNPADTGHGTVIELSGSDGP
CKIPIVSVASLNDMTPVGLTVNPFVATSSANSKVLVEMEPFGDSYIVVGRGDKQI
NHHMKAGSTLTKAFSTTLKGAORLAALGDTAWDFGSGGVFNSIGKAVHQVFGGAFR
TLFGGMSWITQGLMGALLMGMGNARDRSIALAFLATGGLVFLATNVHA"

ORIGIN
Query Match      54.3%; Score 22.8; DB 10; Length 1500;
Best Local Similarity 79.4%; Pred. No. 88;
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 CACGCTTGCCGAGCCCTTACAAAGACTATGAAG 37
    ||||| ||| ||||| ||||| ||||| |||||
Db 1203 CACGCTGGGCAAGCCTTTTCAACGACTTTGAAG 1236

RESULT 40
AY243843 Japanese encephalitis virus isolate A2 polyprotein gene, partial
LOCUS AY243843 1500 bp RNA linear VRL 27-APR-2003
DEFINITION cds.
ACCESSION AY243843
VERSION AY243843.1 GI:30143745
KEYWORDS
SOURCE
ORGANISM Japanese encephalitis virus
Japanese encephalitis virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Flavivirus; Japanese encephalitis virus group.
REFERENCE 1 (bases 1 to 1500)

```

```

/notes="from Walter Reed Army Institute of Research"
<1..>1500
/codon_start=1
/product="envelope protein"
/protein_id="AAD24538.1"
/db_xref="GI:4580916"
/translation="FNCILGMGNRDFIEGASGATWVDLVLEGDSCLTLMANDKPTLDVR
MINIEAVQLAEVRSYCYHASVTDISTVARCPTTGAHNKRRADSSVYCKQGFTRDGMG
NCGGLFGKSIDTCAKFSCTKAIGRTIQPENIKYEVGIFVHGTTTSENHGNYSQVG
ASQAAKFTVTNAPSITLKLGDYGEVTLDCPRSLNTEAFYVMTVSGSKFLVHREWF
HDALPFTPPSTAWNRRELLMESEAHATKQSVVALGQEGGLHQALAGAIIVVEYSN
SVKLTSGHLKCRKMDKALKGTTVMCTEKFSPAKNPADTGHGTVIELSGSDGP
CKIPIVSVASLNDMTPVGLTVNPFVATSSANSKVLVEMEPFGDSYIVVGRGDKQI
NHHMKAGSTLTKAFSTTLKGAORLAALGDTAWDFGSGGVFNSIGKAVHQVFGGAFR
TLFGGMSWITQGLMGALLMGMGNARDRSIALAFLATGGLVFLATNVHA"

ORIGIN
Query Match      54.3%; Score 22.8; DB 10; Length 1500;
Best Local Similarity 79.4%; Pred. No. 88;
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 CACGCTTGCCGAGCCCTTACAAAGACTATGAAG 37
    ||||| ||| ||||| ||||| ||||| |||||
Db 1203 CACGCTGGGCAAGCCTTTTCAACGACTTTGAAG 1236

RESULT 40
AY243843 Japanese encephalitis virus isolate A2 polyprotein gene, partial
LOCUS AY243843 1500 bp RNA linear VRL 27-APR-2003
DEFINITION cds.
ACCESSION AY243843
VERSION AY243843.1 GI:30143745
KEYWORDS
SOURCE
ORGANISM Japanese encephalitis virus
Japanese encephalitis virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Flavivirus; Japanese encephalitis virus group.
REFERENCE 1 (bases 1 to 1500)

```

Li, G., Li, X. and Fu, S.  
 Submitted (25-FEB-2003) State Key Lab. of Molecular Virology &  
 Genetic Engineering, Inst. of Virology, 100 Xingxin St., Xuan Wu  
 Qu, Beijing 100052, China

FEATURES  
 source 1..1500  
 /organism="Japanese encephalitis virus"  
 /mol\_type="genomic RNA"  
 /isolate="A2"  
 /isolation\_source="patient brain sample"  
 /db\_xref="taxon:11072"  
 /country="China"  
 <1..>1500  
 /note="envelope protein region"  
 /codon\_start=1  
 /product="polyprotein"  
 /protein\_id="AAPI4893.1"  
 /db\_xref="GI:30143746"  
 /translation="FNCGLGNRDFIEGASGATWVDLVLEGSCLTIMANDKPTLDR  
 MNIEAQLAEVSYCHASVTDISTVARCPTTGEAHNEKRAADSSVCKQGTDRMG  
 NGGLFGKSIDTCAKFSCTKAIGRTIQSENIKYEVGIFVHGTTTSENHNGYSAQV  
 ASQAAKFTVPNPASTITLKLGDYGEVTLDCPEPSGLNTEAFYVMTVGSKSLVHREWF  
 HDLALPWPSPSTAMNRELLMEFEAHATKQSVVALGSOEGLHQAALAGAIIVVEYSN  
 SVKLTSCHLKRCLKMDKLTGTYTGMCTEKSFAKNPADTGHGTVTIYELISVSGSDGP  
 CKPIVSVASLNDMTPVGRLVTVNPFVATSSANSKVLVEMEPFGDSYIVVGRGDKQI  
 NHHWKAGSTLKGAFSTTLKGAORLAALGDTANDFGSIGVFNISGKAVHQVFGGAFR  
 TLFGGMSWITQGLMGALLWMGNARDRSIALAFLATGGVLVFLATNVHA"

CDS  
 Query Match 54.3%; Score 22.8; DB 10; Length 1500;  
 Best Local Similarity 79.4%; Pred. No. 88;  
 Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

ORIGIN  
 Qy 4 CACGCTTGGCGAGCCCTTACAAAGACTATGAAG 37  
 ||||| ||| |||| |||| |||| |||| |||| |||| ||||  
 Db 1203 CACGCTGGCAAGCCTTTTCAACGACTTTGAAG 1236  
 ||||| ||| |||| |||| |||| |||| |||| |||| ||||  
 RESULT 42  
 JEU70389  
 LOCUS Japanese encephalitis virus mRNA linear VRL 14-JAN-1997  
 DEFINITION Japanese encephalitis virus envelope protein mRNA, partial cds.  
 ACCESSION U70389  
 VERSION U70389.1 GI:1778225  
 KEYWORDS  
 SOURCE Japanese encephalitis virus  
 ORGANISM Japanese encephalitis virus  
 Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 Flavivirus; Japanese encephalitis virus group.  
 REFERENCE 1 (bases 1 to 1500)  
 AUTHORS Chin, C.C., Sung, S.C., Shih, C.M. and Chang, G.J.  
 TITLE Correlation of evolution and phenotype of Japanese encephalitis  
 viruses  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 1500)  
 AUTHORS Chang, G.J.  
 TITLE Direct Submission  
 JOURNAL Submitted (11-SEP-1996) DVID/NCID/CDC/US Department of Health and  
 Human Services, Arbovirus Disease Branch, P.O. Box 2087, Fort  
 Collins, CO 80522, USA  
 FEATURES  
 source Location/Qualifiers  
 1..1500  
 /organism="Japanese encephalitis virus"  
 /mol\_type="mRNA"  
 /strain="Beijing 1/China/1949/Human"  
 /isolate="substrain from DVID/NCID/CDC"  
 /db\_xref="taxon:11072"  
 <1..>1500  
 /function="structural protein"  
 /codon\_start=1  
 /product="envelope protein"  
 /protein\_id="AAB40664.1"  
 /db\_xref="GI:1778226"  
 CDS  
 Query Match 54.3%; Score 22.8; DB 10; Length 1500;  
 Best Local Similarity 79.4%; Pred. No. 88;  
 Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

ORIGIN  
 Qy 4 CACGCTTGGCGAGCCCTTACAAAGACTATGAAG 37  
 ||||| ||| |||| |||| |||| |||| |||| |||| ||||  
 Db 1203 CACGCTGGCAAGCCTTTTCAACGACTTTGAAG 1236  
 ||||| ||| |||| |||| |||| |||| |||| |||| ||||  
 RESULT 41  
 JEU44974  
 LOCUS Japanese encephalitis virus RNA linear VRL 01-AUG-1996  
 DEFINITION Japanese encephalitis virus TL envelope protein gene, partial cds.  
 ACCESSION U44974  
 VERSION U44974.1 GI:1477521  
 KEYWORDS  
 SOURCE Japanese encephalitis virus  
 ORGANISM Japanese encephalitis virus  
 Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 Flavivirus; Japanese encephalitis virus group.  
 REFERENCE 1 (sites)  
 AUTHORS Sung, S.  
 TITLE Genetic variations of envelope protein gene of Japanese  
 encephalitis viruses from Taiwan  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 1500)  
 AUTHORS Sung, S.  
 TITLE Direct Submission  
 JOURNAL Submitted (02-JAN-1996) Shu-Chiun Sung, Division of Molecular  
 Virology, National Defense Medical Center, Taipei, Taiwan, Republic  
 of China  
 FEATURES  
 source Location/Qualifiers  
 1..1500  
 /organism="Japanese encephalitis virus"  
 /mol\_type="genomic RNA"  
 /strain="TL"  
 /db\_xref="taxon:11072"  
 /lab\_host="Aedes albopictus"  
 <1..>1500  
 /function="structural protein"  
 /product="envelope protein"  
 /protein\_id="AAB05394.1"  
 /codon\_start=1

```
Db      1203 CACGCTGGCGAAGCCTTTTCAACGACTTTGAAG 1236

RESULT 43
JEU70413
LOCUS      1500 bp      mRNA      linear      VRL 14-JAN-1997
DEFINITION Japanese encephalitis virus envelope protein mRNA, partial cds.
ACCESSION U70413
VERSION    U70413.1 GI:1778273
KEYWORDS   .
SOURCE     Japanese encephalitis virus
ORGANISM   Japanese encephalitis virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Flavivirus; Japanese encephalitis virus group.
REFERENCE  1 (bases 1 to 1500)
            Chin,C.C., Sung,S.C., Shih,C.M. and Chang,G.J.
            Correlation of evolution and phenotype of Japanese encephalitis
            viruses
JOURNAL    Unpublished
AUTHORS    2 (bases 1 to 1500)
            Chang,G.J.
TITLE      Direct Submission
JOURNAL    Submitted (11-SEP-1996) DVBID/NCID/CDC/US Department of Health and
            Human Services, Arbovirus Disease Branch, P.O. Box 2087, Fort
            Collins, CO 80522, USA
FEATURES   Location/Qualifiers
            source          1..1500
                        /organism="Japanese encephalitis virus"
                        /mol_type="mRNA"
                        /strain="Sagiyama/1957/Japan/Mosquito"
                        /db_xref="taxon:11072"
                        <!--1500
                        /function="structural protein"
                        /codon_start=1
                        /product="envelope protein"
                        /protein_id="AAB40694.1"
                        /db_xref="GI:1778286"
                        /translation="FNCILGMGNRDFIEGASGATWDLVLEGDSCLTMANDKPTLDVR
                        MINIASOLAEVRSYCHASVTDISTVARCPTTGEAHNEKRDSSVYCKOGFTDRGNG
                        NCGLFGKGSIDTCARFSCSTKAIGRTIQPENIKYEVGIFVHGTTTSENHGNYSQVGF
                        ASQAARFTVPNAPSITLKLGDYGEVTLDCPRSGLNTFAFYMTVGSXFLVHREWF
                        HDLALPWPSPSTANRNRELLMEFEAHATKOSAVALGSOEGGLHOALAGAIIVVEYSN
                        SVKLTSGHLKRLMDKALKGTTCMCTEKSFAKNPADTGHGTGVVIELSYSGSDGP
                        CKIPIVSASLNDMTVPGLVTVPFVATSSANSKVLVEMEPFGDSYIVGRGDKQI
                        NHHWHKAGSTLGAFTTLKGAQRLAALGDTAWDFSGIGVFNISGKAVHQVFGGAFR
                        TLFGGMSWITQGLMGALLMGMGNARDRSIALAFLATGGVLVFLATNVHA"
            CDS
            ORIGIN
            Query Match      54.3%; Score 22.8; DB 10; Length 1500;
            Best Local Similarity 79.4%; Pred. No. 88;
            Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY      4 CACGCTTGCCGAGCCCTTACAAGACTATGAAG 37
        ||||| ||| ||||| ||||| ||||| ||||| |||||
Db      1203 CACGCTGGCGAAGCCTTTTCAACGACTTTGAAG 1236

RESULT 45
JEU70420
LOCUS      1500 bp      mRNA      linear      VRL 14-JAN-1997
DEFINITION Japanese encephalitis virus envelope protein mRNA, partial cds.
ACCESSION U70420
VERSION    U70420.1 GI:1778287
KEYWORDS   .
SOURCE     Japanese encephalitis virus
ORGANISM   Japanese encephalitis virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Flavivirus; Japanese encephalitis virus group.
REFERENCE  1 (bases 1 to 1500)
            Chin,C.C., Sung,S.C., Shih,C.M. and Chang,G.J.
            Correlation of evolution and phenotype of Japanese encephalitis
            viruses
JOURNAL    Unpublished
AUTHORS    2 (bases 1 to 1500)
            Chang,G.J.
TITLE      Direct Submission
JOURNAL    Submitted (11-SEP-1996) DVBID/NCID/CDC/US Department of Health and
            Human Services, Arbovirus Disease Branch, P.O. Box 2087, Fort
            Collins, CO 80522, USA
FEATURES   Location/Qualifiers
            source          1..1500
                        /organism="Japanese encephalitis virus"
                        /mol_type="mRNA"
                        /strain="VN118/1979/Vietnam/Mosquito"
                        /db_xref="taxon:11072"
                        <!--1500
                        /function="structural protein"
                        /codon_start=1
                        /product="envelope protein"
                        /protein_id="AAB40695.1"
                        /db_xref="GI:1778288"
            CDS
            ORIGIN
            Query Match      54.3%; Score 22.8; DB 10; Length 1500;
            Best Local Similarity 79.4%; Pred. No. 88;
            Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY      4 CACGCTTGCCGAGCCCTTACAAGACTATGAAG 37
        ||||| ||| ||||| ||||| ||||| ||||| |||||
Db      1203 CACGCTGGCGAAGCCTTTTCAACGACTTTGAAG 1236

RESULT 44
JEU70419
LOCUS      1500 bp      mRNA      linear      VRL 14-JAN-1997
DEFINITION Japanese encephalitis virus envelope protein mRNA, partial cds.
ACCESSION U70419
VERSION    U70419.1 GI:1778285
KEYWORDS   .
SOURCE     Japanese encephalitis virus
ORGANISM   Japanese encephalitis virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Flavivirus; Japanese encephalitis virus group.
REFERENCE  1 (bases 1 to 1500)
            Chin,C.C., Sung,S.C., Shih,C.M. and Chang,G.J.
            Correlation of evolution and phenotype of Japanese encephalitis
            viruses
```

/translation="ENCILGMGNRDFIEGASGATWVDLVLGDSCLTLMADDKPTLDVR  
MINIEASQAEVRSVCYHASVTDISTVARCPTTGBAHNEKRADSSYVCKQGTDRGNG  
NGCGLFGKSIDTCAKFSKAIGRTIOPENIKYEVGIFVHGTTTSENHGNYSQVQ  
ASQAAKFTVTNPAPSTLKLGDYGEVTLDCPSRLNGTEAFYVMTVGRSLVHREWF  
HDALPLPTPPSTAWNRRELLMEFEAHATKQSVVALGSEQGLHQALAGALVVEYS  
SVKLTSGHLKCRLLKMDLALQKTTVGMCTEKFSPAKNPADTGHGTVWIELSVSGSDGP  
CKIPIVSVASLNDMTVPVRLVTVNPFVATSSANSKVLVEMEPFPGDSYIVVGRGDKQI  
NHHWHKAGSTLGKAFSTTLKGQRLLAALGDTAWDFGSGGGFNSIGKAVHQVFGGAFR  
TLFGMSWITQGLMGLLLMGMGNARDRSIALAFIATGGLVFLATNVHA"

Query Match 54.3%; Score 22.8; DB 10; Length 1500;  
Best Local Similarity 79.4%; Pred. No. 88;  
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 4 CACGCTGGCGAGCCCTTACAAAGACTATGAAG 37  
Db 1203 CACGCTGGCGAAAGCCTTTTCAACGACTTTGAAG 1236

RESULT 46  
JVEVEG4  
LOCUS Japanese encephalitis virus (691004) envelope gene.  
DEFINITION  
ACCESSION J34097.1 GI:496912  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Japanese encephalitis virus  
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
Flavivirus; Japanese encephalitis virus group.  
REFERENCE  
1 (bases 1 to 1500)  
Paranjpe, S.G.  
JOURNAL Thesis (1994) Biochemistry, National Institute of Virology  
2 (bases 1 to 1500)  
Paranjpe, S.G.  
AUTHORS  
TITLE Direct Submission  
JOURNAL Submitted (30-MAY-1994) Paranjpe S. G., National Institute of  
Virology, Biochemistry, 20-A, Dr. Ambedkar Road, Pune, Maharashtra,  
India, 411001

FEATURES  
source  
1..1500  
/organism="Japanese encephalitis virus"  
/mol\_type="genomic RNA"  
/strain="691004"  
/isolate="Human blood"  
/db\_xref="taxon:11072"  
/clone="Sri Lanka, 1969"  
1..>1500  
/gene="envelope gene"  
1..>1500  
/gene="envelope gene"  
/function="structural protein of the virus"  
/protein\_id="CAA83983.1"  
/db\_xref="GI:496913"  
/db\_xref="GOA:Q82932"  
/db\_xref="UniProtKB/TrEMBL:Q82932"  
/translation="ENCILGMGNRDFIEGASGATWVDLVLGDSCLTLMADDKPTLDVR  
MINIEAVOLTEVRSVCYHASVTDISTVARCPTTGBAHNEKRADSSYVCKQGTDRGNG  
NGCGLFGKSIDTCAKFSKAIGRTIOPENIKYEVGIFVHGTTTSENHWNYSQVQ  
ASQAAKFTVTNPAPSTLKLGDYGEVTLDCPSRLNGTEAFYVMTVGRSLVHREWF  
RDALPLPTPPSTAWNRRELLMEFEAHATKQSVVALGSEQGLHQALAGALVVEYS  
SVKLTSGHLKCRLLKMDLALQKTTVGMCTEKFSPAKNPADTGHGTVWIELSVSGSDGP  
CKIPIVSVASLNDMTVPVRLVTVNPFVATSSANSKVLVEMEPFPGDSYIVVGRGDKQI  
NHHWHKAGSTLGKAFSTTLKGQRLLAALGDTAWDFGSGGGFNSIGKAVHQVFGGAFR  
TLFGMSWITQGLMGLLLMGMGNARDRSIALAFIATGGLVFLATNVHA"

Query Match 54.3%; Score 22.8; DB 10; Length 1500;  
Best Local Similarity 79.4%; Pred. No. 88;  
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 4 CACGCTGGCGAGCCCTTACAAAGACTATGAAG 37  
Db 1203 CACGCTGGCGAAAGCCTTTTCAACGACTTTGAAG 1236

RESULT 47  
JVEVEG4  
LOCUS Japanese encephalitis virus strain Sarawak polyprotein pseudogene,  
partial sequence.  
DEFINITION  
ACCESSION AY184214.1 GI:30959386  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Japanese encephalitis virus  
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
Flavivirus; Japanese encephalitis virus group.  
REFERENCE  
1 (bases 1 to 1896)  
Solomon, T., Ni, H., Beasley, D.W., Ekkelenkamp, M., Cardosa, M.J. and  
Barrett, A.D.  
JOURNAL Origin and evolution of Japanese encephalitis virus in southeast  
Asia  
2 (bases 1 to 1896)  
Solomon, T., Cardosa, M.J. and Barrett, A.  
AUTHORS Direct Submission  
TITLE Submitted (19-NOV-2002) Neurological Science, Walton Center for  
Neurology and Neurosurgery, University of Liverpool, Lower Lane,  
Fazakerley, Liverpool L9 7LJ, UK

FEATURES  
source  
1..1896  
/organism="Japanese encephalitis virus"  
/mol\_type="genomic RNA"  
/strain="Sarawak"  
/specific\_host="mosquito"  
/db\_xref="taxon:11072"  
/country="Malaysia"  
/note="Isolated in 1965"  
1..>1896  
/note="polyprotein; contains structural proteins"  
/pseudo  
/codon\_start=1

Query Match 54.3%; Score 22.8; DB 10; Length 1896;  
Best Local Similarity 79.4%; Pred. No. 95;  
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 4 CACGCTGGCGAGCCCTTACAAAGACTATGAAG 37  
Db 1839 CACGCTGGCGAAAGCCTTTTCAACGACTTTGAAG 1872

RESULT 48  
BC105147  
LOCUS Bos taurus similar to G protein-coupled receptor, family C, group  
5, member A, mRNA (cdna clone MGC:127231 IMAGE:7945740), complete  
cds.  
DEFINITION  
ACCESSION BC105147  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Bos taurus (cattle)  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;  
Pecora; Bovidae; Bovinae; Bos.  
1 (bases 1 to 2069)  
Moore, S., Alexander, L., Brownstein, M., Guan, L., Lobo, S., Meng, Y.,  
Taniguchi, M., Wang, Z., Yu, J., Prange, C., Schreiber, K., Shenmen, C.,  
Wagner, L., Bala, M., Barbazuk, S., Barber, S., Babakaiff, R.,

Query Match 54.3%; Score 22.8; DB 10; Length 1500;  
Best Local Similarity 79.4%; Pred. No. 88;  
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 4 CACGCTGGCGAGCCCTTACAAAGACTATGAAG 37  
Db 1203 CACGCTGGCGAAAGCCTTTTCAACGACTTTGAAG 1236

RESULT 47  
AY184214  
LOCUS Japanese encephalitis virus strain Sarawak polyprotein pseudogene,  
partial sequence.  
DEFINITION  
ACCESSION AY184214  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Japanese encephalitis virus  
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
Flavivirus; Japanese encephalitis virus group.  
REFERENCE  
1 (bases 1 to 1896)  
Solomon, T., Ni, H., Beasley, D.W., Ekkelenkamp, M., Cardosa, M.J. and  
Barrett, A.D.  
JOURNAL Origin and evolution of Japanese encephalitis virus in southeast  
Asia  
2 (bases 1 to 1896)  
Solomon, T., Cardosa, M.J. and Barrett, A.  
AUTHORS Direct Submission  
TITLE Submitted (19-NOV-2002) Neurological Science, Walton Center for  
Neurology and Neurosurgery, University of Liverpool, Lower Lane,  
Fazakerley, Liverpool L9 7LJ, UK

FEATURES  
source  
1..1896  
/organism="Japanese encephalitis virus"  
/mol\_type="genomic RNA"  
/strain="Sarawak"  
/specific\_host="mosquito"  
/db\_xref="taxon:11072"  
/country="Malaysia"  
/note="Isolated in 1965"  
1..>1896  
/note="polyprotein; contains structural proteins"  
/pseudo  
/codon\_start=1

Query Match 54.3%; Score 22.8; DB 10; Length 1896;  
Best Local Similarity 79.4%; Pred. No. 95;  
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 4 CACGCTGGCGAGCCCTTACAAAGACTATGAAG 37  
Db 1839 CACGCTGGCGAAAGCCTTTTCAACGACTTTGAAG 1872

RESULT 48  
BC105147  
LOCUS Bos taurus similar to G protein-coupled receptor, family C, group  
5, member A, mRNA (cdna clone MGC:127231 IMAGE:7945740), complete  
cds.  
DEFINITION  
ACCESSION BC105147  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Bos taurus (cattle)  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;  
Pecora; Bovidae; Bovinae; Bos.  
1 (bases 1 to 2069)  
Moore, S., Alexander, L., Brownstein, M., Guan, L., Lobo, S., Meng, Y.,  
Taniguchi, M., Wang, Z., Yu, J., Prange, C., Schreiber, K., Shenmen, C.,  
Wagner, L., Bala, M., Barbazuk, S., Barber, S., Babakaiff, R.,



**TITLE**  
JOURNAL

**REMARK**  
COMMENT

Beland, J., Chun, E., Del Rio, L., Gibson, S., Hanson, R., Kirkpatrick, R., Liu, J., Matsuo, C., Mayo, M., Santos, R.R., Stott, J., Tsai, M., Wong, D., Siddiqui, A., Holt, R., Jones, S.J. and Marra, M.A.  
Direct Submission  
Submitted (15-SEP-2005) BC Cancer Agency, Canada's Michael Smith Genome Sciences Centre, Suite 100, 570 West 7th Avenue, Vancouver, British Columbia V5Z 4S6, Canada  
NTH-MGC Project  
Contact: Robert Kirkpatrick  
Canada's Michael Smith Genome Sciences Centre  
BC Cancer Agency  
Suite 100, 570 West 7th Avenue, Vancouver, British Columbia, Canada, V5Z 4S6  
Tel: 1-604-707-5900 x5406  
Fax: 1-604-876-3561  
Email: robertk@bcsc.ca  
Tissue Procurement: M. Taniguchi, Y. Meng, L. Guan and S. Moore, University of Alberta  
cDNA Library Preparation: M. Masaaki, Y. Meng, S. Lobo, L. Guan and Dr. S. Moore  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Bovine Genome Sequencing Program, Genome Sequence Centre,  
BC Cancer Agency, Vancouver, BC, Canada  
info@bcsc.bc.ca  
Moore S, Alexander L, Brownstein M, Guan L, Lobo S, Meng Y, Taniguchi M, Wang Z, Prange C, Schreiber K, Shennen C, Wagner L, Ali J, Bala M, Barber S, Babakoff R, Beland J, Chand S, Chun E, Del Rio L, Gibson S, Kirkpatrick R, Liu J, Matsuo C, Mayo M, Palmquist D, Santos RR, Stott J, Tsai M, Wong D, Wynhoven B, Hanson R, Siddiqui A, Holt R, Jones SJ, Marra MA

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium website at: <http://image.llnl.gov>  
Series: IRAK Plate: 240 Row: n Column: 22.

# FEATURES

**source**

1. .2069  
/organism="Bos taurus"  
/mol\_type="mRNA"  
/strain="Crossbred x Angus"  
/db\_xref="taxon:9913"  
/clone="MGC:127231 IMAGE:7945740"  
/tissue\_type="ileum, crossbred x Angus female calf, 6 months old."  
/clone\_lib="CC\_BGC-02"  
/lab\_host="XL10-Gold"  
/notes="Vector: pCMV-SPORT6.0"  
1. .2069  
/genes="MGC127231"  
/db\_xref="GeneID:516026"  
136..1209  
/genes="MGC127231"  
/codon\_start=1  
/product="hypothetical protein LOC516026"  
/protein\_id="AAI05148.1"  
/db\_xref="GI:75773418"  
/db\_xref="GeneID:516026"  
/translations="MSTVAPGRSLDSRYRLCKKAAGIVLEALAGVANASVA FMITLLVICKVDNSKRLKLLPQFLIGLVGIFGLTAFITDIDGGTGPRFLFG VLFALCFGLVHAFNLKLVGRQPLSMVLMGLGFLSLVQDIIAIEYVLTWNRT NVNIFSELSPRRNEDFVMLLIYVLFMALTLTASFSFGSFTAWKRHGAHLYTTI LSAIIVWMTLLVPTLSPQWDDTLSSALVANGWVLLAVIAPFHLITRQONPMD YPVEDAFCPQPKQSYGVNRYASQEGIIQGTTEGTSILYAPSTHFLQNRDSKD FSIPIRQTRVSPYSYDEGRKDV5"

# ORIGIN

**Query Match** 54.3%; Score 22.8; DB 14; Length 2069;  
**Best Local Similarity** 71.4%; Pred. No. 98;  
**Matches** 30; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

**Qy** 1 GCCACGCTTGCCGCGCCCTTACAAGACTATGAAGTAAAG 42  
|||||  
1153 GTCCAGACTCGGGTTAGCCCTTACGTGACTACGAAGGAAG 1194

**Db**

# RESULT 49

S75726  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

S75726  
Japanese encephalitis virus polyprotein gene, partial cds.  
S75726  
S75726.1 GI:913060

Japanese encephalitis virus  
Viruses; sRNA positive-strand viruses, no DNA stage; Flaviviridae;  
Flavivirus; Japanese encephalitis virus group.  
1 (bases 1 to 2379)  
Hasegawa, H., Yoshida, M., Fujita, S. and Kobayashi, Y.  
Comparison of structural proteins among antigenically different  
Japanese encephalitis virus strains  
Vaccine 12 (9), 841-844 (1994)  
7975862  
GenBank  
entry [NCBI gi359161325] from the original journal article.

# FEATURES

**source**

1. .2379  
/organism="Japanese encephalitis virus"  
/mol\_type="genomic RNA"  
/db\_xref="taxon:11072"  
/note="Nakayama-RFVL"  
<1..>2379  
/note="conceptual translation presented here differs from translation in publication"  
/codon\_start=1  
/product="polyprotein"  
/protein\_id="AA832599.2"  
/db\_xref="GI:9864807"  
/translation="TKKPGGPKGNRAINMKRLPRVFLVGVKRVMSLLDGRGPVR  
KRLGATITFKALPTALGRKAVKSVAMKHLTSFKRELGTLDIVVNRKGRKON  
KRGNEGSIIMLASLAVIACAGAMKLSNFQGLKLTNNVDIADIVIVPNSKGRNC  
WYRAIDVGYMCEDTITYECPKLTMGNDPDDVDCWQNOEVVYQYGRCTLRHSKRSR  
SVSVQTHGESSLVNKEAWLDSKTATRYLMKTNWIVNPGVAFALAILGWLMSNG  
QRVFTLLLVAPAYSFNCLGMGRDFEGASGATWDLVLEGDSCLTIMANDKPTL  
DYRMINIEAVQLAEVSYCHASVTDISTVARCPTTGEAHNEKRDSSVYCKQGFTR  
GWNCGCLFGKSIDTCFKFSCTKAIGRTIQPENIKYEVGIFVHGTTTSENHGYSA  
QVGAOAAKFTVTPNAPSITLKLGDYGEVLTDCPRSLNTEAFVMTVGSKSLVHR  
EFNFDLALPWTPTSTANRELLMEFEAHATKOSVALGSOEGLHOLAAGAIIVE  
YSSVKLTSGHLKCRKMDKLAIKGTTCMCTEKSFAKNPADTGHGTVIELSYSGS  
DGCKIPIVSVASLNDMPVGRLVTPNPFVATSSANKVLVMEPEPFQDSIIVVGRGD  
KQINQHHKAGSTLTKAFSTTLKGAQLAALGDTAWDFSGISGVFNISKAVHQVFGG  
APRTLFGGMSWITQGLMGALLLMGMVGNARDRSIALAFLATGRVLFVFLATNVHA"

# CDS

**Qy** 4 CACGCTTGCGCGAGCCCTTACAAGACTATGAAG 37  
|||||  
2082 CACGCTGGCGCAAGCCCTTTTCAACGACTTTGAAG 2115

**Db**

# ORIGIN

**Query Match** 54.3%; Score 22.8; DB 10; Length 2379;  
**Best Local Similarity** 79.4%; Pred. No. 1e+02;  
**Matches** 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

# RESULT 50

JEU03696  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

JEU03696  
Japanese encephalitis virus Saigon polyprotein gene, partial cds.  
U03696  
U03696.1 GI:517409

Japanese encephalitis virus  
Viruses; sRNA positive-strand viruses, no DNA stage; Flaviviridae;  
Flavivirus; Japanese encephalitis virus group.  
1 (bases 1 to 2436)  
Ni, H. and Barrett, A.D.  
Nucleotide and deduced amino acid sequence of the structural

protein genes of Japanese encephalitis viruses from different geographical locations  
J. Gen. Virol. 76 (Pt 2), 401-407 (1995)  
7844559  
2 (bases 1 to 2436)  
Ni, H.  
Direct Submission  
Submitted (23-NOV-1993) Haolin Ni, Pathology, University of Texas Medical Branch, F-05, Galveston, TX 77555-0066, USA  
Location/Qualifiers  
1. .2436  
/organism="Japanese encephalitis virus"  
/mol\_type="genomic RNA"  
/strain="Saigon"  
/db\_xref="taxon:11072"  
/clone="pFRL9"  
misc\_feature  
1. .2436  
/note="this sequence corresponds to nucleotides 28-2463 in the full JE virus genome"  
69. .>2436  
/note="encodes C, prM, M, and E proteins"  
/codon\_start=1  
/product="polyprotein"  
/protein\_id="AA67165.1"  
/db\_xref="GI:517410"  
/translation="MKKPGGPKRINMLKRGLPVRVPLVGVKRVVMSLLDGRGPV RFLALITFFKALAPKALGRWAKVEKSVAMKHLTSFKRELGLTDLAVNKRGRQ NKRGNESIMWLASLAVIACAGAKLSNFQGLLMTVNNTDIAVTAIDTSKGNR CWRAIDVYMCELDITCEPKLGMNDPEDVDCWNOEVQVGRCTRNRHRSKR RSVSVQTHGESSLVNKEAWLDSTOATRYLAKMTENIIVRNPGYAFLLAVGLWMLGSNN GQRVFIILLLVAPAYFNCGLIGNRDFIEGASGATWDLVLESDSLTIMANDKPT LDVRMINLEAVRSYCYHASVTIDISTVARCPTTEAHNEKRADSSYVCKQGFDT PWNGCGCLFGKSGIDTCAKFSCTRAIGRTIQENIKYEVGIFVHGTTTSENHNYS AQVGASQAAKFTVPNDPSTLLKLDYGEVTLDCPRSGLNTAEFYVMTVGSKPLVH REWFHDLPLTSPSTNRNRELMFEGAHATKQSVVAGLSQEGGLHQALAGAIV EYSSVKLTSHLKRCLKMDKLAKGTTGYMCTEKFSKPNADPGHGTGVVIELSYSG SDGCPKPIVSVASLNDMTPVGRLVTVNPFVATSSVNSKVLVEMEPFGDSYIVVGRG DQKINHHKAGSTGTFSTLKGQRLAALGDTAWDFGSIIGGVFNFSIGKAIHOVFG GAFRTLFGGMSWITQGLMGLLWMGVNARDRSIALAFLVTGGVLFLA"  
ORIGIN  
Query Match 54.3%; Score 22.8; DB 10; Length 2436;  
Best Local Similarity 79.4%; Pred. No. 1e+02;  
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
Qy 4 CACGCTTGCCGAGCCCTTACAAAGACTATGAAG 37  
|||||  
Db 2153 CACGCTGGCAAGCCCTTTCAACGACTTTGAAG 2186  
|||||  
RESULT 51  
JJEVPR  
LOCUS  
DEFINITION  
Japanese encephalitis virus capsid protein (3' end), membrane protein, envelope protein, nonstructural proteins NS1 and NS2, complete cds and nonstructural protein NS3 (3' end) mRNA.  
M16574.1 GI:331336  
M16574  
4320 bp ss-RNA linear VRL 02-AUG-1993  
Japanese encephalitis virus  
protein, envelope protein, nonstructural proteins NS1 and NS2, complete cds and nonstructural protein NS3 (3' end) mRNA.  
SOURCE  
ORGANISM  
Japanese encephalitis virus  
Virus; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Flavivirus; Japanese encephalitis virus group.  
REFERENCE  
1 (bases 1 to 4320)  
McAda, P.C., Mason, P.W., Schmaljohn, C.S., Dalrymple, J.M., Mason, T.L. and Fournier, M.J.  
Partial nucleotide sequence of the Japanese encephalitis virus genome  
Virology 158 (2), 348-360 (1987)  
JOURNAL  
PUBMED  
3035787  
COMMENT  
Original source text: Japanese encephalitis virus (strain Nakaya), CDNA to viral RNA, passed 54 times in suckling mice.

Draft entry and computer-readable sequence for [1] kindly provided by D.Taylor, 29-SEP-1987.  
Location/Qualifiers  
1. .4320  
/organism="Japanese encephalitis virus"  
/mol\_type="genomic RNA"  
/db\_xref="taxon:11072"  
<1. .>4320  
/note="polyprotein"  
/codon\_start=1  
/protein\_id="AAA46251.1"  
/db\_xref="GI:331337"  
/translation="SVAMKHLTSFKRELGLTDLAVNKRGRKONKRGNGESIMWLASL AVIACAGAKLSNFQGLLMTVNNTDIAVIVPNPSKGENRCWRAIDVYMCEDT IYCEPKLTGMNDPEDVDCWNOEVQVGRCTRHSKRSRSVSQTHGESSLVN KKEAWLDSTKATRYLMTKNTIIVRNPGYAFLLAVGLWMLGNGNQRRWYFTLLILVA PAYFNCGLMGNRDFIEGASGATWDLVLEGSDSLTIMANDKPTLDVRMINLEAVQLA EVRSYCVHASVTIDISTVARCPTTEAHNEKRADSSYVCKQGFDRGWNGCGLFGKGS IDTCAKFSCTSKAIGRTIQENIKYEVGIFVHGTTTSENHNYSQVGSQAQAKTIVT NPAISITLKLGDYGEVTLDCPRSGLNTAEFYVMTVGSKFLVHREWFHDLALPWTTPP SSTAWNRNRELMFEFEAHATKQSVVAGLSQEGGLHQALAGAIIVSEYSSVKLTSGHLK CRLKMDKLAKGTTGYMCTEKFSKPNADPGHGTGVIELSYSGSDGCKPIVSVAS LNDMTPVGRLVTVNPFVATSSANSKVLVEMEPFGDSYIVVGRGSKQINHHWKAGST LGKAFSTTLKGAORLAALGDTAWDFGSIIGGVFNFSIGKAVHGVFGCAFRTLFGEMCSG OGLGALLWMGVNARDRSIALAFLATGGVLAVELATNVHADTGCAIDITRKEMRCSG IIVHNDPAWDRYKILPETPRSLAKIVHKAHKEGVCGRSVTRLEHQWNEAVRDELN VLLKENAVDLSVNVNKNPVGRSAPKRLSMTQEKPEMGWAKGKILFAPELANSTFV VDGTEPCDPEDHAWNSIEIDDFGGITSTRVWLKIREESTDECDAIIGTAVKGHV AVHSDLSYIESRVNDTWKRAVGEVKSCTPETHTLWGDGVESESLIIPHTIAGP KSKHNRGTYQNGQWPDENIGVLDYFCPTKVITTEDCGKRGPSVRTTIDSGKLI TDWCRCSSLPRLPERTENGCMYGMERIPVRHDETLTVRSOVDAENGWMDVPOLGLL WMLQATVLRKRWARTLTIIPAVIGALLVLMGGITYTDLARYVVLVAAAFAEASGG DVHLALIAVFKIQIAPFLVMNMLSTRNTQNVNVLVGAAPFLHASVDLQIGVHGLN AAAATAMIVRAITPTTSSVTPVLLATPGMRALYLDTRYIILVIIGICSLQBRKK TMAKKGVALILLALTTSTGFSPTTIAAGLMVCNPKRGWPAFTFLSAVGLMFAIVG GAELDIESMSIPMLAGLMAVSYVSGKATDMLERAADISWENDAAITGSSRRLDV KLDDGDGPHLLDDPGVPKWKVLMRMSCIGLAAITPAIAPAAFGYWLTKTKTKRGGVF WDTIP"  
mat\_peptide  
159  
/product="capsid protein"  
sig\_peptide  
160. .438  
/note="membrane protein signal peptide"  
mat\_peptide  
439. .666  
/product="membrane protein"  
mat\_peptide  
667. .2166  
/product="envelope protein"  
mat\_peptide  
2167. .3408  
/product="nonstructural protein NS1"  
mat\_peptide  
3409. .3903  
/product="nonstructural protein ns2a"  
mat\_peptide  
3904. .4296  
/product="nonstructural protein ns2b"  
mat\_peptide  
4297. .>4320  
/product="nonstructural protein NS3"  
ORIGIN 1908 bp upstream of SacI site.  
Query Match 54.3%; Score 22.8; DB 10; Length 4320;  
Best Local Similarity 79.4%; Pred. No. 1.3e+02;  
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
Qy 4 CACGCTTGCCGAGCCCTTACAAAGACTATGAAG 37  
|||||  
Db 1869 CACGCTGGCAAGCCCTTTCAACGACTTTGAAG 1902  
|||||  
RESULT 52  
AR003118  
LOCUS  
DEFINITION  
Sequence 52 from patent US 5744140.  
AR003118  
ACCESSION  
AR003118.1 GI:3964377  
VERSION  
KEYWORDS  
Unknown.



Best Local Similarity 79.4%; Pred. No. 1.3e+02; Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;		/db_xref="taxon:32630"	
QY	4	CACGCTGGCGAGCCCTTACAAAGACTATGAAG 37 	

```

JOURNAL      Patent: JP 2001299336-A 2 30-OCT-2001;
COMMENT      EIJI KONISHI
              OS Artificial Sequence
              PN JP 2001299336-A/2
              PD 30-OCT-2001
              PP 28-APR-2000 JP 2000130250
              PI EIJI KONISHI, PETER W MASON
              PC C12N5/10.A61K39/12.A61P31/12.C07K14/18.C12N15/09.C12P21/02//
              PC (C12N5/10.C12R1.91).C12N5/00.C12N15/00.(C12N5/00.C12R1.91) CC
              Preparation of Flavivirus-like grains that are stable and non-CC
              toxic as
              CC vaccine and diagnostic antigen
              FH Key Location/Qualifiers
              FT source 1..7486
              FT /organism='Artificial Sequence'.
              FT Location/Qualifiers
              FEATURES
                source
                1..7486
                /organism='synthetic construct'
                /mol_type='genomic DNA'
                /db_xref='taxon:32630'
ORIGIN
Query Match 54.3%; Score 22.8; DB 2; Length 7486;
Best Local Similarity 79.4%; Pred. No. 1.5e+02;
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 CACGCTGGCGAGCCCTTACAAGACTATGAAG 37
    ||||| ||| ||||| ||||| ||||| ||||| |||||
Db 2710 CACGCTGGCGAAGCCTTTTCAACGACTTTGAAG 2743

RESULT 59
LOCUS      BD064746 7486 bp DNA linear PAT 27-AUG-2002
DEFINITION Expression vector containing cDNA encoding Japanese encephalitis
              virus and vaccine using it.
ACCESSION BD064746
VERSION 1 GI:22610349
KEYWORDS JP 2001299355-A/2.
SOURCE synthetic construct
ORGANISM other sequences; artificial sequences.
REFERENCE 1 (bases 1 to 7486)
AUTHORS Mason, P.W., Konishi, E., Yamaoka, M. and Kurane, I.
TITLE Expression vector containing cDNA encoding Japanese encephalitis
              virus and vaccine using it
JOURNAL Patent: JP 2001299355-A 2 30-OCT-2001;
              DIRECTOR GENERAL OF NATIONAL INSTITUTE OF INFECTIONS DISEASES,
              ICHIRO KURANE, EIJI KONISHI
COMMENT OS Artificial Sequence
              PN JP 2001299355-A/2
              PD 30-OCT-2001
              PP 27-APR-2000 JP 2000126989
              PI PETER W MASON, EIJI KONISHI, MASAKI YAMAOKA, ICHIRO KURANE PC
              C12N15/09.A61K39/12.A61P31/12.A61P31/12./C07K14/18.C12P21/02.PC
              (C12N15/09.C12R1.92).C12N15/00.(C12N15/00.C12R1.92) CC Expression
              vector containing cDNA encoding Japanese CC
              encephalitis virus and
              CC vaccine using it
              FH Key Location/Qualifiers
              FT source 1..7486
              FT /organism='Artificial Sequence'.
              FT Location/Qualifiers
              FEATURES
                source
                1..7486
                /organism='synthetic construct'
                /mol_type='genomic DNA'
                /db_xref='taxon:32630'
ORIGIN
Query Match 54.3%; Score 22.8; DB 2; Length 7486;
Best Local Similarity 79.4%; Pred. No. 1.5e+02;
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

JOURNAL      Patent: JP 2001299336-A 2 30-OCT-2001;
COMMENT      EIJI KONISHI
              OS Artificial Sequence
              PN JP 2001299336-A/2
              PD 30-OCT-2001
              PP 28-APR-2000 JP 2000130250
              PI EIJI KONISHI, PETER W MASON
              PC C12N5/10.A61K39/12.A61P31/12.C07K14/18.C12N15/09.C12P21/02//
              PC (C12N5/10.C12R1.91).C12N5/00.C12N15/00.(C12N5/00.C12R1.91) CC
              Preparation of Flavivirus-like grains that are stable and non-CC
              toxic as
              CC vaccine and diagnostic antigen
              FH Key Location/Qualifiers
              FT source 1..7486
              FT /organism='Artificial Sequence'.
              FT Location/Qualifiers
              FEATURES
                source
                1..7486
                /organism='synthetic construct'
                /mol_type='genomic DNA'
                /db_xref='taxon:32630'
ORIGIN
Query Match 54.3%; Score 22.8; DB 2; Length 7486;
Best Local Similarity 79.4%; Pred. No. 1.5e+02;
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 CACGCTGGCGAGCCCTTACAAGACTATGAAG 37
    ||||| ||| ||||| ||||| ||||| ||||| |||||
Db 2710 CACGCTGGCGAAGCCTTTTCAACGACTTTGAAG 2743

RESULT 60
LOCUS      JEVBEICG 10976 bp ss-RNA linear VRL 21-NOV-1995
DEFINITION Japanese encephalitis virus (strain Beijing-1), complete genome.
ACCESSION L48961
VERSION 1 GI:1066797
KEYWORDS Japanese encephalitis virus
SOURCE Japanese encephalitis virus
ORGANISM Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
              Flavivirus; Japanese encephalitis virus group.
REFERENCE 1 (sites)
AUTHORS Hashimoto, H., Nomoto, A., Watanabe, K., Mori, T., Takezawa, T.,
              Aizawa, C., Takegami, T. and Hiramatsu, K.
TITLE Molecular cloning and complete nucleotide sequence of the genome of
              Japanese encephalitis virus Beijing-1 strain
JOURNAL Virus Genes 1 (3), 305-317 (1988)
PUBMED 3245133
COMMENT Original source text: Japanese encephalitis virus (strain
              Beijing-1) RNA.
FEATURES
                source
                1..10976
                Location/Qualifiers
                /organism='Japanese encephalitis virus'
                /mol_type='genomic RNA'
                /strain='Beijing-1'
                /db_xref='taxon:11072'
ORIGIN
Query Match 54.3%; Score 22.8; DB 10; Length 10976;
Best Local Similarity 79.4%; Pred. No. 1.8e+02;
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 CACGCTGGCGAGCCCTTACAAGACTATGAAG 37
    ||||| ||| ||||| ||||| ||||| ||||| |||||
Db 2180 CACGCTGGCGAAGCCTTTTCAACGACTTTGAAG 2213

RESULT 61
LOCUS      AC025723/c 77422 bp DNA linear INV 07-SEP-2005
DEFINITION Caenorhabditis elegans cosmid Y54F10AM, complete sequence.
ACCESSION AC025723
VERSION AC025723.2 GI:13559726
KEYWORDS HTG.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
              Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
              Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 77422)
AUTHORS C. elegans Sequencing Consortium
CONSTRM Genome sequence of the nematode C. elegans: a platform for
              investigating biology
TITLE Science 282 (5396), 2012-2018 (1998)
JOURNAL PUBMED 9851916
REFERENCE 2 (bases 1 to 77422)
AUTHORS Ryan, E., Courtney, L. and Yoakum, M.
TITLE The sequence of C. elegans cosmid Y54F10AM
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 77422)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (12-MAR-2000) Genome Sequencing Center, Washington
              University School of Medicine, 4444 Forest Park Parkway, St. Louis,
              MO 63108, USA
REFERENCE 4 (bases 1 to 77422)
AUTHORS Waterston, R.
TITLE Direct Submission

```









-----

#### NOTICE:

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

#### SOURCE INFORMATION:

The RPCI-24 BAC Library has been constructed by Pieter de Jong and coworkers (<http://www.chori.org>) from male C57BL/6J mouse spleen and/or brain genomic DNA. The clone and detailed information can be obtained from Pieter de Jong and coworkers at <http://www.chori.org>

#### NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone.

#### FEATURES

source	Location/Qualifiers
repeat_region	1. .145213
repeat_region	/organism="Mus musculus"
repeat_region	/mol_type="Genomic DNA"
repeat_region	/db_xref="taxon:10090"
repeat_region	/chromosomes="1"
repeat_region	/map="1"
repeat_region	/clone="RP24-173118"
repeat_region	/clone_lib="RPCI-24"
repeat_region	4. .2142
repeat_region	/rpt_family="L1"
repeat_region	7798..8187
repeat_region	/rpt_family="ERVK"
repeat_region	8188..8662
repeat_region	/rpt_family="L1"
repeat_region	8667..10718
repeat_region	/rpt_family="L1"
repeat_region	12405..12534
repeat_region	/rpt_family="Alu"
repeat_region	13132..13246
repeat_region	/rpt_family="Alu"
repeat_region	13441..13507
repeat_region	/rpt_family="ID"
repeat_region	13742..13883
repeat_region	/rpt_family="Alu"
repeat_region	13948..14263
repeat_region	/rpt_family="MaLR"
repeat_region	14264..14437
repeat_region	/rpt_family="ERVL"
repeat_region	14576..14702
repeat_region	/rpt_family="Alu"
repeat_region	15146..15952
repeat_region	/rpt_family="L1"
repeat_region	15953..16137
repeat_region	/rpt_family="B2"
repeat_region	16138..16688
repeat_region	/rpt_family="L1"
repeat_region	16686..19681
repeat_region	/rpt_family="L1"
repeat_region	21390..21519
repeat_region	/rpt_family="Alu"
repeat_region	22118..22210
repeat_region	/rpt_family="Alu"
repeat_region	22418..22488
repeat_region	/rpt_family="ID"
repeat_region	22683..22824
repeat_region	/rpt_family="Alu"
repeat_region	22891..23382
repeat_region	/rpt_family="MaLR"
repeat_region	23497..23632
repeat_region	/rpt_family="Alu"
repeat_region	24137..24255
repeat_region	/rpt_family="Alu"
repeat_region	24275..24403
repeat_region	/rpt_family="Alu"
repeat_region	24412..24662
repeat_region	/rpt_family="L1"
repeat_region	24698..24807
repeat_region	/rpt_family="B4"
repeat_region	25032..25187
repeat_region	/rpt_family="L1"
repeat_region	25192..25633
repeat_region	/rpt_family="L1"
repeat_region	25751..26167
repeat_region	/rpt_family="ERVK"
repeat_region	26168..26350
repeat_region	/rpt_family="B2"
repeat_region	26351..26480
repeat_region	/rpt_family="ERVK"
repeat_region	26508..26655
repeat_region	/rpt_family="Alu"
repeat_region	26656..26999
repeat_region	/rpt_family="ERVK"
repeat_region	27000..27315
repeat_region	/rpt_family="MaLR"
repeat_region	27531..27934
repeat_region	/rpt_family="ERVK"
repeat_region	28312..28365
repeat_region	/rpt_family="MaLR"
repeat_region	28372..28563
repeat_region	/rpt_family="B2"
repeat_region	28564..28705
repeat_region	/rpt_family="B4"
repeat_region	29337..29460
repeat_region	/rpt_family="B4"
repeat_region	29506..30142
repeat_region	/rpt_family="L1"
repeat_region	30245..30307
repeat_region	/rpt_family="4.5SRNA"
repeat_region	30260..30327
repeat_region	/rpt_family="Alu"
repeat_region	31197..31340
repeat_region	/rpt_family="Alu"
repeat_region	33345..33443
repeat_region	/rpt_family="Alu"
repeat_region	34066..34180
repeat_region	/rpt_family="Alu"
repeat_region	34374..34429
repeat_region	/rpt_family="ID"
repeat_region	34670..34811
repeat_region	/rpt_family="Alu"
repeat_region	34876..35267
repeat_region	/rpt_family="MaLR"
repeat_region	35268..35365
repeat_region	/rpt_family="ERVL"
repeat_region	35483..35627
repeat_region	/rpt_family="Alu"
repeat_region	36082..36881
repeat_region	/rpt_family="L1"
repeat_region	36882..37067
repeat_region	/rpt_family="B2"
repeat_region	37068..37618
repeat_region	/rpt_family="L1"
repeat_region	37616..38978
repeat_region	/rpt_family="L1"
repeat_region	38982..39335
repeat_region	/rpt_family="ERVK"
repeat_region	39392..39574
repeat_region	/rpt_family="ERVK"
repeat_region	39575..40131

```

/rpt_family="ERVK"
40132..40409
/rpt_family="ERVK"
40409..41158
/rpt_family="ERVK"
41161..41587
/rpt_family="ERVK"
41588..43274
/rpt_family="L1"
43660..44187
/rpt_family="L1"

Query Match      54.3%; Score 22.8; DB 6; Length 145213;
Best Local Similarity 71.4%; Pred. No. 4.3e+02;
Matches 30; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1 GCCACGCTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42
Db 28332 GCCACGATGGCTGAGCTTTAAAAAATAAAAAAAG 28291

RESULT 64
AC095186/c
LOCUS          247690 bp      DNA      linear      HTG 13-NOV-2002
DEFINITION    Rattus norvegicus clone CH230-9B21, WORKING DRAFT SEQUENCE, 3
ACCESSION     AC095186
VERSION       HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS      Rattus norvegicus (Norway rat)
SOURCE        Rattus norvegicus
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Murioidea; Muridae; Murinae; Rattus.
REFERENCE     Muzny,D,Marie, Metaker,M, Lee, Abranzon,S, Adams,C, Alder,J.,
Allen,C, Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregregis,B., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hognes,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuheva,L., Loulasegh,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokemele,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasteriak,S., Paul,H., Perez,A., Perez,L., Pfanckoch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puzo,M., Quirroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,

```

```

Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
Sheed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Staimie,M., Strong,R., Sutton,A., Svatek,A., Taber,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Umani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wlarczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 247690)
Worley,K.C.
Direct Submission
Submitted (16-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 247690)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (13-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 13, 2002 this sequence version replaced gi:22772986.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GDBT
Center clone name: CH230-9B21
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 215114 bases at least Q40
Consensus quality: 217886 bases at least Q30
Consensus quality: 219758 bases at least Q20
Estimated insert size: 221396; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 101341: contig of 101341 bp in length
* 101342 101441: gap of unknown length
* 219685 219685: contig of 118244 bp in length
* 219686 219785: gap of unknown length
* 219786 247690: contig of 27905 bp in length.
Location/Qualifiers
1..247690
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
FEATURES
source

```

```

/db xref="taxon:10116"
/clone="CH230-9821"
misc_feature
1..1604
/notes="wgs_contig"
gap
101342..101441
/estimated_length=unknown
219686..219785
/estimated_length=unknown

ORIGIN
Query Match      54.3%; Score 22.8; DB 12; Length 247690;
Best Local Similarity 79.4%; Pred. No. 5.2e+02;
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 9 TTGCGCGAGCCCTTACAAAGACTATGAGACTAAAG 42
Db 200484 TAGGACCAACACTTTCAAGAGACTTTGAAGTAAAG 200451

RESULT 65
AC137204
LOCUS
DEFINITION
AC137204.1 GI:25073092 linear HTG 19-NOV-2002
Rattus norvegicus clone CH230-unknown, *** SEQUENCING IN PROGRESS
***, 4 unordered pieces.
ACCESSION
AC137204.1 GI:25073092
VERSION
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
KEYWORDS
Rattus norvegicus (Norway rat)
SOURCE
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Rattus.
1 (bases 1 to 249669)

REFERENCE
AUTHORS
Muzny, D., Marie, E., Metzker, M., Lee, S., Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebregiorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogue, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenshewa, L., Loulseghe, H., Lozado, R.J., Lu, X., Ma, J.,
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwaokelimeh, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,
Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reich, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, R., Ruiz, S.J.,
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,

```

```

Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tincey, A., Trejos, Z., Umami, K.,
Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
Weinstock, G. and Gibbs, R.A.
Direct Submission
Unpublished
2 (bases 1 to 249669)
Rat Genome Sequencing Consortium.
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: K2PB
Center clone name: CH230-unknown
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 20161 bases at least Q40
Consensus quality: 206093 bases at least Q30
Consensus quality: 208911 bases at least Q20
Estimated insert size: 209892; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 103590: contig of 103590 bp in length
* 103591 103690: gap of unknown length
* 103691 218904: contig of 115214 bp in length
* 218905 219004: gap of unknown length
* 219005 248412: contig of 29408 bp in length
* 248413 248512: gap of unknown length
* 248513 249669: contig of 1157 bp in length.
* Location/Qualifiers
1..249669
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-unknown"
1..1761
/notes="wgs_contig"
4671..6761
/notes="wgs_contig"
103591..103690
FEATURES
source
misc_feature
misc_feature
gap

```

```

/estimated_length=unknown
218905. .219004
/estimated_length=unknown
248413. .248512
/estimated_length=unknown

ORIGIN
Query Match          54.3%; Score 22.8; DB 12; Length 249669;
Best Local Similarity 79.4%; Pred. No. 5.2e+02;
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 9 TTGGCGGACCTTACAAGACTATGACGTAAG 42
Db 123020 TAGGACCAACACTTTCAAAGACTTTGAAGTAAAG 123053

RESULT 66
AC006875      298216 bp      DNA      linear      HTG 24-FEB-1999
Caenorhabditis elegans clone Y54F10x, *** SEQUENCING IN PROGRESS
***, 25 unordered pieces.
AC006875
AC006875.1 GI:4263473
HTG; HTGS PHASE1
Caenorhabditis elegans
Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 298216)
Waterston.R.H.
The sequence of Caenorhabditis elegans clone
Unpublished
2 (bases 1 to 298216)
Waterston.R.H.
Direct Submission
Submitted (24-FEB-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
* NOTE: This is a 'working draft' sequence. It currently
* consists of 25 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 2000: contig of 2000 bp in length
* 2001 2015: gap of unknown length
* 2016 3995: contig of 1980 bp in length
* 3996 4010: gap of unknown length
* 4011 7604: contig of 3594 bp in length
* 7605 7619: gap of unknown length
* 7620 10290: contig of 2671 bp in length
* 10291 10305: gap of unknown length
* 10306 14792: contig of 4487 bp in length
* 14793 14807: gap of unknown length
* 14808 18212: contig of 3405 bp in length
* 18213 18227: gap of unknown length
* 18228 21270: contig of 3043 bp in length
* 21271 21285: gap of unknown length
* 21286 24062: contig of 2777 bp in length
* 24063 24077: gap of unknown length
* 24078 27102: gap of unknown length
* 27088 27102: gap of unknown length
* 27103 32662: contig of 5560 bp in length
* 32663 32677: gap of unknown length
* 32678 38716: contig of 6039 bp in length
* 38717 38731: gap of unknown length
* 38732 42014: contig of 3282 bp in length
* 42014 42028: gap of unknown length
* 42029 47980: contig of 5952 bp in length
* 47981 47995: gap of unknown length
* 47996 52552: contig of 4557 bp in length

/estimated_length=unknown
218905. .219004
/estimated_length=unknown
248413. .248512
/estimated_length=unknown

FEATURES             source
1.298216
/organism="Caenorhabditis elegans"
/mol_type="genomic DNA"
/db_xref="taxon:6239"
/clone="Y54F10x"
2001. .2015
/estimated_length=unknown
3996. .4010
/estimated_length=unknown
7605. .7619
/estimated_length=unknown
10291. .10305
/estimated_length=unknown
14793. .14807
/estimated_length=unknown
18213. .18227
/estimated_length=unknown
21271. .21285
/estimated_length=unknown
24083. .24077
/estimated_length=unknown
27088. .27102
/estimated_length=unknown
32663. .32677
/estimated_length=unknown
38717. .38731
/estimated_length=unknown
42014. .42028
/estimated_length=unknown
47981. .47995
/estimated_length=unknown
52553. .52567
/estimated_length=unknown
59007. .59021
/estimated_length=unknown
66562. .66576
/estimated_length=unknown
73618. .73632
/estimated_length=unknown
84893. .84907
/estimated_length=unknown
97462. .97476
/estimated_length=unknown
110333. .110347
/estimated_length=unknown
128092. .128106
/estimated_length=unknown
166559. .166573
/estimated_length=unknown
201396. .201410
/estimated_length=unknown
52553 52567: gap of unknown length
52568 59006: contig of 6439 bp in length
59007 59021: gap of unknown length
66561 66562: contig of 7540 bp in length
66562 66567: gap of unknown length
73617 73632: contig of 7041 bp in length
73632 84892: gap of unknown length
84892 84907: contig of 11260 bp in length
84907 97461: contig of 12554 bp in length
97461 97476: gap of unknown length
97476 110332: contig of 12856 bp in length
110332 110347: gap of unknown length
110347 128091: contig of 17744 bp in length
128091 128106: gap of unknown length
128106 166558: contig of 38452 bp in length
166558 166573: gap of unknown length
166573 201395: contig of 34822 bp in length
201395 20410: gap of unknown length
20410 275281: contig of 73871 bp in length
275281 275296: gap of unknown length
275296 298216: contig of 22920 bp in length.
Location/Qualifiers
1.298216
/organism="Caenorhabditis elegans"
/mol_type="genomic DNA"
/db_xref="taxon:6239"
/clone="Y54F10x"
2001. .2015
/estimated_length=unknown
3996. .4010
/estimated_length=unknown
7605. .7619
/estimated_length=unknown
10291. .10305
/estimated_length=unknown
14793. .14807
/estimated_length=unknown
18213. .18227
/estimated_length=unknown
21271. .21285
/estimated_length=unknown
24083. .24077
/estimated_length=unknown
27088. .27102
/estimated_length=unknown
32663. .32677
/estimated_length=unknown
38717. .38731
/estimated_length=unknown
42014. .42028
/estimated_length=unknown
47981. .47995
/estimated_length=unknown
52553. .52567
/estimated_length=unknown
59007. .59021
/estimated_length=unknown
66562. .66576
/estimated_length=unknown
73618. .73632
/estimated_length=unknown
84893. .84907
/estimated_length=unknown
97462. .97476
/estimated_length=unknown
110333. .110347
/estimated_length=unknown
128092. .128106
/estimated_length=unknown
166559. .166573
/estimated_length=unknown
201396. .201410
/estimated_length=unknown

```

```

/estimated_length=unknown
275282..275296
/estimated_length=unknown

Query Match      54.3%   Score 22.8;   DB 12;   Length 298216;
Best Local Similarity 71.4%   Pred. No. 5.5e+02;
Matches 30;   Conservative 0;   Mismatches 12;   Indels 0;   Gaps 0;

ORIGIN
1 GCCACGCTTGGCGGACCCCTTACAAGACTATGAAGTAAAG 42
2 ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 152561 GCAGATGCTTGTTCGAGCCATTGGAAGAGCGCTGAAGGATAG 12602

RESULT 67
AC141041
LOCUS      AC141041      161575 bp      DNA      linear      HTG 27-MAR-2003
DEFINITION Rattus norvegicus clone CH230-473J6, *** SEQUENCING IN PROGRESS
***, 61 unordered pieces.
ACCESSION  AC141041
VERSION    GI:28875900
KEYWORDS  HTG; HTGS_PHASE1.
SOURCE    Rattus norvegicus (Norway rat)
ORGANISM  Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Murioidea; Muridae; Murinae; Rattus.
1 (bases 1 to 161575)
Muzny,D.Marie., Metzker,M., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gegeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guvera,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowig,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuewa,L., Louleghed,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhinney,S., McLeod,M., McNeill,T., Meenen,E., Milosavljevic,A.,
Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K.,
Morris,S., Mundasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D.,
Newton,N., Nguyen,N., Norris,S., Nwaokemele,O., Okwuonu,G.,
Olanrinsunagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H.,
Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A.,
Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E.,
Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y.,
Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A.,
Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S.,
Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A.,
Sisson,I., Sitter,C.D., Smays,D., Sneed,A., Sodergren,E.,
Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A.,
Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S.,
Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D.,
Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J.,
Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R.,
Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S.,

```

```

Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X.,
Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R.,
Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 161575)
Worley,K.C.
Direct Submission
Submitted (07-MAR-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 161575)
Worley,K.C.
Direct Submission
Submitted (27-MAR-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GXIU
Center clone name: CH230-473J6
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 130671 bases at least Q40
Consensus quality: 139673 bases at least Q30
Consensus quality: 145030 bases at least Q20
Estimated insert size: 131321; sum-of-contigs estimation
Quality coverage: 2x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 61 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1276: contig of 1276 bp in length
* 1277 1376: gap of unknown length
* 1377 2670: contig of 1294 bp in length
* 2671 2770: gap of unknown length
* 2771 4049: contig of 1279 bp in length
* 4050 4149: gap of unknown length
* 4150 5234: contig of 1085 bp in length
* 5235 5334: gap of unknown length
* 5335 6376: contig of 1041 bp in length
* 6376 6475: gap of unknown length
* 6476 7950: contig of 1475 bp in length
* 7951 8050: gap of unknown length
* 8051 10029: contig of 1979 bp in length
* 10030 10129: gap of unknown length
* 10130 11297: contig of 1168 bp in length
* 11298 11397: gap of unknown length
* 11398 12571: contig of 1174 bp in length
* 12572 12671: gap of unknown length
* 12672 13872: contig of 1201 bp in length
* 13873 13972: gap of unknown length
* 13973 15359: contig of 1387 bp in length
* 15360 15459: gap of unknown length
* 15460 16729: contig of 1270 bp in length
* 16730 16829: gap of unknown length
* 16830 17925: contig of 1095 bp in length
* 17925 18024: gap of unknown length
* 18025 19369: contig of 1345 bp in length
* 19370 19469: gap of unknown length

```

\* 19470 21151: contig of 1682 bp in length  
\* 21152 21251: gap of unknown length  
\* 21252 22627: contig of 1376 bp in length  
\* 22628 22727: gap of unknown length  
\* 22728 24187: contig of 1460 bp in length  
\* 24188 24287: gap of unknown length  
\* 24288 26783: contig of 2496 bp in length  
\* 26784 26883: gap of unknown length  
\* 26884 27222: contig of 1039 bp in length  
\* 27923 28022: gap of unknown length  
\* 28023 29392: contig of 1270 bp in length  
\* 29393 29392: gap of unknown length  
\* 29393 31112: contig of 1920 bp in length  
\* 31113 31412: gap of unknown length  
\* 31413 32519: contig of 1107 bp in length  
\* 32520 32619: gap of unknown length  
\* 32620 34673: contig of 2054 bp in length  
\* 34674 34773: gap of unknown length  
\* 34774 36319: contig of 1546 bp in length  
\* 36320 36419: gap of unknown length  
\* 36420 37971: contig of 1552 bp in length  
\* 37972 38071: gap of unknown length  
\* 38072 40564: contig of 2493 bp in length  
\* 40565 40664: gap of unknown length  
\* 40665 42016: contig of 1352 bp in length  
\* 42017 42116: gap of unknown length  
\* 42117 43834: contig of 1718 bp in length  
\* 43835 45295: contig of 1361 bp in length  
\* 45296 45395: gap of unknown length  
\* 45396 47677: contig of 2282 bp in length  
\* 47678 47777: gap of unknown length  
\* 47778 49906: contig of 2129 bp in length  
\* 49907 50006: gap of unknown length  
\* 50007 52238: contig of 2232 bp in length  
\* 52239 52338: gap of unknown length  
\* 52339 53800: contig of 1462 bp in length  
\* 53801 53900: gap of unknown length  
\* 53901 55788: contig of 1888 bp in length  
\* 55789 55889: gap of unknown length  
\* 55889 58800: contig of 2912 bp in length  
\* 58801 58900: gap of unknown length  
\* 58901 62150: contig of 3250 bp in length  
\* 62151 62250: gap of unknown length  
\* 62251 64203: contig of 1953 bp in length  
\* 64204 64303: gap of unknown length  
\* 64304 67105: contig of 2802 bp in length  
\* 67106 67205: gap of unknown length  
\* 67206 69322: contig of 2727 bp in length  
\* 69323 70032: gap of unknown length  
\* 70033 72191: contig of 2159 bp in length  
\* 72192 72291: gap of unknown length  
\* 72292 75506: contig of 3215 bp in length  
\* 75507 75606: gap of unknown length  
\* 75607 77496: contig of 1890 bp in length  
\* 77497 77596: gap of unknown length  
\* 77597 80560: contig of 2964 bp in length  
\* 80561 80660: gap of unknown length  
\* 80661 83785: contig of 3125 bp in length  
\* 83786 83885: gap of unknown length  
\* 83886 86387: contig of 2502 bp in length  
\* 86388 86487: gap of unknown length  
\* 86488 89888: contig of 3401 bp in length  
\* 89889 92744: gap of unknown length  
\* 92745 92844: gap of unknown length  
\* 92845 97427: contig of 4583 bp in length  
\* 97428 97527: gap of unknown length  
\* 97528 100290: contig of 2763 bp in length  
\* 100291 100390: gap of unknown length

Query Match 53.8%; Score 22.6; DB 12; Length 161575;  
Best Local Similarity 75.7%; Pred. No. 5.5e+02;

Matches 28; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
Qy 5 ACGCTGGCCGAGCCCTTACAAAGACTATGAAGTAA 41  
Db 21098 AGGTTTGGCCGAGCCCTTTCAATTCCTATAAAGGAA 21134  
RESULT 68  
AC166012/c  
LOCUS  
DEFINITION Oryctolagus cuniculus clone LB1-138P6, WORKING DRAFT SEQUENCE, 5  
ACCESSION AC166012  
VERSION AC166012.2 GI:71980158  
KEYWORDS HTG; HTGS PHASE2; HTGS DRAFT.  
SOURCE Oryctolagus cuniculus (rabbit)  
ORGANISM Oryctolagus cuniculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Lagomorpha;  
Leporidae; Oryctolagus.  
REFERENCE 1 (bases 1 to 173071)  
AUTHORS Antonellis,A., Ayele,K., Bass,D., Benjamin,B., Bera,J.,  
Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S., Chu,G.,  
Coleman,H., Engle,J., Franks,S., Fuksenko,T., Gestole,M.,  
Greene,A., Guan,X., Gupta,J., Gurson,N., Haghighi,P., Han,J.,  
Hansen,N., Ho,S.-L., Hu,P., Hunter,G., Hurie,B., Idol,J.R.,  
Kwong,P., Laric,P., Larson,S., Lee-tin,S.-Q., Legaspi,R.,  
Madden,M., Maduro,Q.L., Maduro,V.B., Margulies,E.H., Mastello,C.,  
Maskeri,B., McDowell,J., Mojidi,H.A., Mullikin,J.C., Park,M.,  
Portnoy,M.E., Prasad,A., Puri,O., Rantz,K., Reddix-Dugue,N.,  
Sante,A., Schandler,K., Schueler,M.G., Sison,C., Stantripop,S.,  
Taye,A., Thomas,J.W., Thomas,P.J., Tsipouri,V., Ung,L., Vogt,J.L.,  
Wetherby,K.D., Withers,T.R., Young,A. and Green,E.D.  
NISC Comparative Sequencing Initiative  
UNPUBLISHED  
REFERENCE 2 (bases 1 to 173071)  
AUTHORS Green,E.D.  
TITLE Direct Submission  
JOURNAL Submitted (20-JUL-2005) NIH Intramural Sequencing Center, 5625  
Fishers Lane, Rockville, MD 20852, USA  
REFERENCE 3 (bases 1 to 173071)  
AUTHORS Green,E.D.  
TITLE Direct Submission  
JOURNAL Submitted (09-AUG-2005) NIH Intramural Sequencing Center, 5625  
Fishers Lane, Rockville, MD 20852, USA  
COMMENT On Aug 9, 2005 this sequence version replaced gi:71037466.  
----- Genome Center  
Center: NIH Intramural Sequencing Center  
Center code: NISC  
Web site: <http://www.nisc.nih.gov>  
Contact: [nisc\\_zoo@nhgri.nih.gov](mailto:nisc_zoo@nhgri.nih.gov)  
----- Project Information  
Center project name: keo  
Center clone name: 138P06

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is generally based on at least 8x average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics  
Sequencing vector: plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 172156 bases at least Q40  
Consensus quality: 172470 bases at least Q30



\* 22784 22883: gap of unknown length  
\* 22884 25416: contig of 2533 bp in length  
\* 25417 25116: gap of unknown length  
\* 25117 28197: contig of 2681 bp in length  
\* 28198 28297: gap of unknown length  
\* 28298 30593: contig of 2236 bp in length  
\* 30594 30693: gap of unknown length  
\* 30694 33550: contig of 2857 bp in length  
\* 33551 37481: contig of 3831 bp in length  
\* 37482 37581: gap of unknown length  
\* 37582 39009: contig of 2328 bp in length  
\* 39010 40009: gap of unknown length  
\* 40010 42913: contig of 2904 bp in length  
\* 42914 43013: gap of unknown length  
\* 43014 46368: contig of 3355 bp in length  
\* 46369 46468: gap of unknown length  
\* 46469 51176: contig of 4708 bp in length  
\* 51177 51276: gap of unknown length  
\* 51277 54607: contig of 3331 bp in length  
\* 54608 54707: gap of unknown length  
\* 54708 59400: contig of 4693 bp in length  
\* 59401 64255: contig of 4755 bp in length  
\* 64256 64356: gap of unknown length  
\* 64357 69910: contig of 5555 bp in length  
\* 69911 70010: gap of unknown length  
\* 70011 75918: contig of 5908 bp in length  
\* 75919 76018: gap of unknown length  
\* 76020 83100: contig of 7082 bp in length  
\* 83101 83200: gap of unknown length  
\* 83201 87479: contig of 4279 bp in length  
\* 87480 87579: gap of unknown length  
\* 87580 93930: contig of 6351 bp in length  
\* 93931 94030: gap of unknown length  
\* 94031 99775: contig of 5745 bp in length  
\* 99776 99875: gap of unknown length  
\* 99876 105199: contig of 5324 bp in length  
\* 105200 105299: gap of unknown length  
\* 105300 111500: contig of 6201 bp in length  
\* 111501 111600: gap of unknown length  
\* 111601 120914: contig of 9314 bp in length  
\* 120915 121014: gap of unknown length  
\* 121015 128821: contig of 7807 bp in length  
\* 128822 128921: gap of unknown length  
\* 128922 138027: contig of 9106 bp in length  
\* 138028 138127: gap of unknown length  
\* 138128 147606: contig of 9479 bp in length  
\* 147607 147706: gap of unknown length  
\* 147707 153904: contig of 6198 bp in length  
\* 153905 154004: gap of unknown length  
\* 154005 164023: contig of 10019 bp in length  
\* 164024 164123: gap of unknown length  
\* 164124 174952: contig of 10829 bp in length  
\* 174953 175052: gap of unknown length  
\* 175053 188622: contig of 13570 bp in length.

FEATURES  
source

1. .188622  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="RP11-693B19"  
/clone="RP11-693B19"  
1. .1660  
/note="assembly\_name:Contig10"  
1661.1760  
/estimated\_length=unknown  
1761. .2764  
/note="assembly\_name:Contig11"  
2765. .2864  
/estimated\_length=unknown  
2865. .4001  
/note="assembly\_name:Contig17"

misc\_feature

gap

misc\_feature

gap

misc\_feature

gap 4002. .4101  
/estimated\_length=unknown  
misc\_feature 4102. .6019  
/note="assembly\_name:Contig18"  
6020. 6119  
/estimated\_length=unknown  
misc\_feature 6120. .7712  
/note="assembly\_name:Contig19"  
7713. .7812  
/estimated\_length=unknown  
misc\_feature 7813. 9594  
/note="assembly\_name:Contig20"  
9595. .9694  
/estimated\_length=unknown  
misc\_feature 9695. .11105  
/note="assembly\_name:Contig21"  
11106. .11205  
/estimated\_length=unknown  
misc\_feature 11206. .12957  
/note="assembly\_name:Contig22"  
12958. .13057  
/estimated\_length=unknown  
misc\_feature 13058. .15321  
/note="assembly\_name:Contig23"  
15322. .15421  
/estimated\_length=unknown  
misc\_feature 15422. .17637  
/note="assembly\_name:Contig25"  
17638. .17737  
/estimated\_length=unknown  
misc\_feature 17738. .20175  
/note="assembly\_name:Contig26"  
20176. .20275  
/estimated\_length=unknown  
misc\_feature 20276. .22783  
/note="assembly\_name:Contig27"  
22784. .22883  
/estimated\_length=unknown  
misc\_feature 22884. .25416  
/note="assembly\_name:Contig28"  
25417. .25516  
/estimated\_length=unknown  
misc\_feature 25517. .28197  
/note="assembly\_name:Contig29"  
28198. .28297  
/estimated\_length=unknown  
misc\_feature 28298. .30593  
/note="assembly\_name:Contig30"  
30594. 30693  
/estimated\_length=unknown  
misc\_feature 30694. .33550  
/note="assembly\_name:Contig31"

Query Match 53.8%; Score 22.6; DB 12; Length 188622;  
Best Local Similarity 75.7%; Pred No. 5.8e+02;  
Matches 28; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 3 CCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTA 39  
|||||  
Db 75903 CCACACCTGGCCAGCCCTTACAAATTTTGTGAATA 75867  
|||||

RESULT 70  
DQ211932/c

LOCUS

DEFINITION

product.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

DQ211932 1857 bp mRNA linear INV 11-OCT-2005  
Toxoplasma gondii mitochondrial alternative NADH dehydrogenase 1  
(anhd-1) mRNA, complete cds; nuclear gene for mitochondrial  
product.  
DQ211932  
DQ211932.1 GI:76561723  
Toxoplasma gondii  
Toxoplasma gondii





Coding sequences below are the result of integration and manual review of the following data : computer analysis using the program GeneFinder (P. Green and L. Hillier, personal communication), the large scale EST projects of Yuji Kohara ([http://www.ddbj.nig.ac.jp/c-elegans/html/CE\\_INDEX.html](http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html)) and The C. elegans ORFeome cloning project (<http://worldb.dfci.harvard.edu/>), similarity to other proteins from Blastx analyses (<http://blast.wustl.edu/>), sequence conservation with C. briggsae using Jim Kent's WABA alignment program (Genome Research 10:1115-1125, 2000), individual C. elegans GenBank submissions, and personal communications with C. elegans researchers. tRNAs are predicted using the program tRNAscan-SE (Lowe, T.M. and Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).

## FEATURES

source

```

1. .11959
/organism="Caenorhabditis elegans"
/mol_type="genomic DNA"
/strain="Bristol N2"
/db_xref="taxon:6239"
/chromosomes="v"
/clone="Y32G9B"
/complement(35. .5738)
/gene="Y32G9B.1"
/locus_tag="Y32G9B.1"
/complement(join(35. .592,1515. .1954,5401. .5465,5617. .5738))
/gene="Y32G9B.1"
/standard_names="Y32G9B.1"
/notes="contains similarity to Ashbya gossypii ABR038Cp.;
TR:Q75D16; coded for by the following C. elegans cDNAs:
YK210f9.3, YK210f9.5"
/codon_start=1
/product="Hypothetical protein Y32G9B.1"
/protein_id="AAK71405.1"
/db_xref="GI:14625315"
/db_xref=" WormBase:Y32G9B.1"
/translation="MDFQFFLSLRPNRSMVAELTSLSSWAGNNMRENRDVQRTV
AAPKPTPEQRSMQIAQNLAFQGRNAFLQALDKLEASGASTPTTPPTPTLP
PLTFGTPTPTPTLPFMFGIPKQESNDNTFDNGRDLGVRDFKKNVAMGEKIE
IRYAEPIAEVAEAPAPVPEIAEAPASASAAVSAPELAPPAPKNAFDGSPF
MQIARRFLQVGTGTPTEGGKAELPQIGRIDFMDNTDSNGLPRGKGLPFTSEFM
QIAYCKQDQADEKTFDQAGELRLAILHGEDLLKASQETCRGAERQCDLSREAV
ATCDVMASLEIGTQLQRAKRCCEVSGMMDQSPAALVQLQLNLTIGEVAQFLNNFLK
NG"

```

## ORIGIN

```

Query Match      53.3%; Score 22.4; DB 13; Length 11959;
Best Local Similarity 81.2%; Pred No. 2.8e+02;
Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

```

QY 1 GCCCAGCTTGCCGAGCCCTTACAAAGACTA 32
||| ||||| ||||| ||||| ||||| |||||
Db 7178 GCTCACCTTGCGCTTAAACAAGACTA 7147

```

## RESULT 72

AC135368/c

LOCUS

```

DEFINITION Homo sapiens X BAC RP13-120P16 (Roswell Park Cancer Institute Human
BAC Library) complete sequence.

```

ACCESSION

AC135368

VERSION

AC135368.2

KEYWORDS

HTG.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.

REFERENCE

1 (bases 1 to 34483)

AUTHORS

Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-ouman, F.R., Allen, C.,

Alsbrooks S.L., Amaratunge, H.C., Are, J.R., Banks, T., Barbara, J.,

Benton, J., Binage, K., Blankenburg, K., Bonnin, D., Bouck, J.,

Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C.,

Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Dunn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgall, D., Edwards, C.C., Elhaj, C., Escott, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseghe, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhinney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, U., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shooshitari, N., Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmali, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Washington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G. and Gibbs, R.

Direct Submission  
Unpublished  
2 (bases 1 to 34483)  
Worley, K.C.  
Direct Submission  
Submitted (13-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 34483)  
Worley, K.C.  
Direct Submission  
Submitted (24-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Oct 24, 2002 this sequence version replaced gi:23928468.  
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email [gc-help@bcm.tmc.edu](mailto:gc-help@bcm.tmc.edu)

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

## ANNOTATION OF FEATURES:

STS are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL:

http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

## FEATURES

## Location/Qualifiers

1. .34483  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="X"  
/clone="RP13-120P16"

misc\_feature 1. .2005

/notes="overlaps bases 119624..121628 of clone AC117517"  
/function="clone overlap"

repeat\_region 127. .200

/rpt\_family="L1PA16"

repeat\_region 466. .486

/rpt\_family="(T)n"

repeat\_region 945. .1025

/rpt\_family="L1PA16"

complement(1288..1571)

/rpt\_family="AluY"

repeat\_region complement(1920..2014)

/rpt\_family="L2"

repeat\_region 2118. .2157

/rpt\_family="(TG)n"

complement(2694..3316)

/rpt\_family="L1ME3A"

repeat\_region 3749. .3895

/rpt\_family="MER58A"

complement(4057..4240)

/rpt\_family="MER58A"

STS 4243. .4491

/standard\_name="STS-T47077"

repeat\_region 4264. .4371

/rpt\_family="L2"

repeat\_region 5068. .5372

/rpt\_family="AluYb8"

repeat\_region 5373. .5395

/rpt\_family="(CAA)n"

repeat\_region 5412. .5563

/rpt\_family="MIR"

repeat\_region 5591. .5614

/rpt\_family="AT\_rich"

repeat\_region 5644. .5716

/rpt\_family="CT-rich"

repeat\_region 5921. .5950

/rpt\_family="AT\_rich"

repeat\_region 6072. .6500

/rpt\_family="MLT1J"

complement(6686..6898)

/rpt\_family="MIR"

repeat\_region 7236. .7282

/rpt\_family="(TCTA)n"

complement(7284..7565)

/rpt\_family="AluX"

repeat\_region complement(7672..8117)

/rpt\_family="MER49"

repeat\_region complement(8174..8381)

/rpt\_family="LTR59"

complement(8420..8719)

/rpt\_family="AluY"

repeat\_region 9116. .9421

repeat\_region /rpt\_family="AluSg"  
complement(9656..10044)  
/rpt\_family="L1"  
10180..10201  
/rpt\_family="AT\_rich"  
complement(10202..10316)  
/rpt\_family="L1MA9"  
complement(10552..10646)  
/rpt\_family="MIR"  
complement(10778..10908)  
/rpt\_family="MIR"  
12280..12301  
/rpt\_family="(T)n"  
complement(12473..12801)  
/rpt\_family="L2"  
complement(12824..13099)  
/rpt\_family="LTR16C"  
complement(13125..13464)  
/rpt\_family="L2"  
13471..13634  
/rpt\_family="MER63A"  
complement(13858..14026)  
/rpt\_family="MIR"  
complement(14033..14097)  
/rpt\_family="L1MC5"  
complement(14116..14302)  
/rpt\_family="L1MC/D"  
14310..15141  
/rpt\_family="L1PA7"  
complement(15182..15328)  
/rpt\_family="L1MB8"  
complement(15385..15432)  
/rpt\_family="L1MC4"  
15521..15696  
/rpt\_family="MER5B"  
16163..16317  
/rpt\_family="MER5A"  
16294..16425  
STS

Query Match 53.3%; Score 22.4; DB 5; Length 34483;  
Best Local Similarity 72.5%; Pred. No. 4e+02;  
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 2 CCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAA 41

Db 12445 CCACACCTGGCATGCCATGACAAATACGATCAAGTTAA 12406

RESULT 73

CRS22870\_21/c

WPCOMMENT

Sequence split into 36 fragments LOCUS CRS22870 Accession CRS22870

Fragment Name	Begin	End
CRS22870_00	1	110000
CRS22870_01	100001	210000
CRS22870_02	200001	310000
CRS22870_03	300001	410000
CRS22870_04	400001	510000
CRS22870_05	500001	610000
CRS22870_06	600001	710000
CRS22870_07	700001	810000
CRS22870_08	800001	910000
CRS22870_09	900001	1010000
CRS22870_10	1000001	1110000
CRS22870_11	1100001	1210000
CRS22870_12	1200001	1310000
CRS22870_13	1300001	1410000
CRS22870_14	1400001	1510000
CRS22870_15	1500001	1610000
CRS22870_16	1600001	1710000
CRS22870_17	1700001	1810000
CRS22870_18	1800001	1910000
CRS22870_19	1900001	2010000

CR522870_20		2000001	2110000
CR522870_21		2100001	2210000
CR522870_22		2200001	2310000
CR522870_23		2300001	2410000
CR522870_24		2400001	2510000
CR522870_25		2500001	2610000
CR522870_26		2600001	2710000
CR522870_27		2700001	2810000
CR522870_28		2800001	2910000
CR522870_29		2900001	3010000
CR522870_30		3000001	3110000
CR522870_31		3100001	3210000
CR522870_32		3200001	3310000
CR522870_33		3300001	3410000
CR522870_34		3400001	3510000
CR522870_35		3500001	3523383
Continuation (22 of 36) of CR522870 from base 2100001 (CR522870 Desulfotalea psychrophila)			
Query Match 53.3%; Score 22.4; DB 15; Length 110000;			
Beat Local Similarity 72.5%; Pred. No. 6+02;			
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;			
Qy	3	CCACGCTTGCGGAGCCCTTACAAGACTATCAAGTAAAG 42	
Db	27621	CCCTGCTTAGCCTGCCATAAAATATTATTAAAGG 27582	
RESULT 74			
PPU344068 116580 bp DNA circular BCT 15-APR-2005			
LOCUS Pseudomonas putida plasmid pW0 complete genome.			
DEFINITION AJ344068			
ACCESSION AJ344068.1 GI:18077099			
VERSION complete genome.			
KEYWORDS Pseudomonas putida			
SOURCE Pseudomonas putida			
ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.			
REFERENCE 1			
AUTHORS Greated,A., Lambertsen,L., Williams,P.A. and Thomas,C.M.			
TITLE Complete sequence of the IncP-9 TOL plasmid pW0 from Pseudomonas putida			
JOURNAL Environ. Microbiol. 4 (12), 856-871 (2002)			
PUBMED 12534468			
REFERENCE 2 (bases 1 to 116580)			
AUTHORS Thomas,C.M.			
TITLE Direct Submission			
JOURNAL Submitted (07-AUG-2001) Thomas C.M., Biosciences, University of Birmingham, Edgbaston, Birmingham, UNITED KINGDOM			
FEATURES			
source			
1..116580			
/organism="Pseudomonas putida"			
/mol_type="genomic DNA"			
/db_xref="taxon:303"			
/plasmid="pW0"			
<1..355			
/gene="parC"			
<1..355			
/gene="parC"			
/codon_start=2			
/transl_table=11			
/protein_id="CAC86736.1"			
/db_xref="GI:18077100"			
/db_xref="UniProtKB/TrEMBL:Q8VMQ3"			
/translation="SLAGINSLRLTLADLRLLETPGPHDRKNWTEKLQAHQKDDQ			
ALFDLHQAGDMGRYPGVDQLVKDHGEDEAKLSPIIRALEHRKVTSAELNELEAAQ			
PILSLRPFPSKLG"			
625..1273			
/gene="res"			
625..631			
/gene="res"			
638..1273			
/gene="res"			

gene	1460..1700	/gene="oriV"
rep_origin	1460..1700	/gene="oriV"
gene	1720..2284	/gene="rep"
RBS	1720..1725	/gene="rep"
CDS	1730..2284	/gene="rep"
/codon_start=1		
/transl_table=11		
/protein_id="CAC86738.1"		
/db_xref="GI:18077102"		
/db_xref="UniProtKB/TrEMBL:Q8VMQ1"		
/translation="MASDNEIRAYAQPAQRGTWVQTRAGAEWAALTAQAPRAAQ		
MHILVOHMKOGALIISOATLAKLMTSVANHKTIRYHSYKHNTOTISVGQRGGL		
AVVNSRLAMADKRDNLQFALFNARLVSTEDQADLGDAKUKQLFTMADGDIQLPAGP		
GMDPQRSLEGLLPDMPSIPHNS"		
RBS	2300..2306	/note="ORF4"
CDS	2313..2606	/note="ORF4"
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
/protein_id="CAC86739.1"		
/db_xref="GI:18077103"		
/db_xref="UniProtKB/TrEMBL:Q8VMQ0"		
/translation="MPOPSNETFEQELHAKAACQYDMNSDYPKGAHSVSEELLGOGET		
RBS	2582..2589	/note="ORF5"
CDS	2596..3105	/note="ORF5"
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
/protein_id="CAC86740.1"		
/db_xref="GI:18077104"		
/db_xref="UniProtKB/TrEMBL:Q8VMP9"		
/translation="MSDERYARLQSLDSAKOHLVDLTGALALPIGSDRDEGISAW		
WQLTGLQLVHNSGLDEATIQELRAIDQLAKATTKEVDQALVASEADGETAAALAD		
FTASHWFKQSLSQALLPRDPVDVAWDAEWLFELLNKRCSVAQLQDVAAEQPMNMEFKAD		
GSTTQIDIT"		
RBS	3266..3271	/note="ORF6"
CDS	3279..4553	/note="ORF6"
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
/protein_id="CAC86741.1"		
/db_xref="GI:18077105"		
/db_xref="UniProtKB/TrEMBL:Q8VMP8"		
/translation="MADQRYAVSIGFGENVMQYSEAYOTQGYALQWKGVLQKAHRVG		
LVRLCPGKGTTRLAVRHRSDSDSFHLSPYPTGAEHALDCLLYSPDPKSLGSGYSK		
GVWETKDGDLKVLTLSLRKDPADVPQMEGTSPTSGGKPTKPMPTLGGLLHLS		
EAGLNHWYPGMAGKRGGLIHSLLQQAADRLASMRIGDALVIATSSAVGQQAANS		
QKVTSAIINNRLVIAFLAAYSEBEQTSRYITITGYHGVPRLTDTDTDTQRIARS		

RBS YTAELRWGREGRRVIAIVOTDPTAPGKAQALNVALMVVSDWIVPESGYEAKTEALL  
RBS REQRRFLKPMRFADAGQSVFDFWLMDDIAGRELPMVEYGRHDPKYLARQAKCAYI  
CDS TOHYCPTGWASWDASADPRGMNIPFPPIRHP"  
4559. .4565  
/note="ORF7"  
4575. .5309  
/note="ORF7"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="CAC86742.1"  
/db\_xref="GI:18077106"  
/db\_xref="GOA:Q8VMP7"  
/db\_xref="UniProtKB/TrEMBL:Q8VMP7"  
/db\_xref="InterPro:IPR005480"  
/translation="MTVNWELVDSKLGKQVDELARLIGSTPHRRRRRREQLGAPWS  
VQKIEPKYHLGYTTDRAIAACQCVSKSVKAYRESLGIIPFVFPVKQALPLNH  
LRPKPLFGYFVDSDEIAQLAGADLAVQDIREALGFEPVPVPGSSPQLIADYHG  
LGYESLSTMTPAKISRHVGVYSIILQCKDLGIQPTTRVSVASRYDHLGKVPD  
VARLAGVSASIGARRQRLGIASATR"  
complement(5405. .6613)  
/gene="ruvB"  
complement(5405. .6613)  
/gene="ruvB"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAC86743.1"  
/db\_xref="GI:18077107"  
/db\_xref="GOA:Q8VMP6"  
/db\_xref="InterPro:IPR001126"  
/db\_xref="UniProtKB/TrEMBL:Q8VMP6"  
/translation="MCSKDRPIFGLVDNSFCYCERIPELRRHPPVVVLSNSDGC  
IVATNEKALGLKMGDPYFKRNFLEQNGVAVSSNTLYGEISHRVGVIAISLVE  
WEQSIDFVNLDELPEFIIDVGRAIQARVLQWALPVGVGIGHTKTLAKAAQHASK  
VMREKTPFLPDPADEWLLRMPVDEVMGIGSRMDNLARDGITTAWLSQADAR  
AMGRKSVLRTRELQGESCMLEQTVAPKTCSSKMFQQRVHTIQLGQALATY  
IHRAAKLRLQSLCGAMRIQIOTSFHGDGPKYANAATCIPAYPTDDVRQLTKVALQ  
AEQIYRPGFAYSKAEVLLDLRKRGEYTQDLFSPSQPERSDALMATMDKINRWGHC  
LRTAAGATDP"  
complement(join(6603. .7034,7039. .7046))  
/gene="ruvA"  
complement(6603. .7034)  
/gene="ruvA"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAC86744.1"  
/db\_xref="GI:18077108"  
/db\_xref="GOA:Q8VMP5"  
/db\_xref="InterPro:IPR006197"  
/db\_xref="UniProtKB/TrEMBL:Q8VMP5"  
/translation="MNVVSIQPLRDSRQVLAVLAGSAGGFPSPAADYIERPPI  
LNIRAAHVFIYRVGESMRDAGIFDGLVIVDRSITPCSGQIVLAYVDNQPLKRLAK  
APGDGWLLESANPAYKTTTPDEYASVDVFGVVTWDLTKHF"  
complement(6625. .6630)  
/gene="ruvB"  
complement(6625. .6630)  
/gene="ruvB"  
complement(7039. .7046)  
/gene="ruvA"  
7158. .7163  
/note="ORF11"  
7172. .8122  
/note="ORF11"  
/codon\_start=1  
/transl\_table=11  
/product="putative ligase"  
/protein\_id="CAC86745.1"  
/db\_xref="GI:18077109"  
/db\_xref="GOA:Q8VMP4"  
/db\_xref="InterPro:IPR000977"  
/translation="MAVKKSQLPARIQPLQVALLTRPAGEWAYEIKFDGYRMFARID

AGVGLFTKNGYDWTARMPLAODLRSLPIRGAWLDGEVTVODVDRPAFOALQAEAFAT  
GRTHLVYFAEDLLFIENVDLRSRVERRELLRVLLBOVDLDQVYSYDLDADPSLL  
LASACALGIEGIVGKRLGSVAGRDGSKWIKLKVCORQEFIILGFTRESSAGISLLIG  
LHNDAGQIQYAGRVRSRGTFRRELDRLKERLPLVRITTFAMEAPPKLSGAVVWAEPLV  
CEVXPAELTPAGKVRHAYVIALREDKPASALSLESDDTPV"  
8142. .8147  
/note="ORF12"  
8153. .8428  
/note="ORF12"  
/codon\_start=1

Query Match 53.3%; Score 22.4; DB 15; Length 116580;  
Best Local Similarity 72.5%; Pred. No. 6.1e+02;  
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1 GCCACGCTTGCCCGAGCCCTTACAAAGACTATGAAGTAA 40  
Db 98830 GCCCGCATGACCGCGCGTGTACCAAGAGCATGGAGAA 98869

RESULT 75  
AC068600/c  
LOCUS AC068600 Homo sapiens BAC clone RP11-567N20 from 2, complete sequence.  
DEFINITION AC068600  
ACCESSION AC068600  
VERSION AC068600.5 GI:15341622  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
REFERENCE 1 (bases 1 to 136657)  
AUTHORS Abbott,A., Spalding,L., Phillips,A., Elliott,G. and Grewal,N.  
TITLE The sequence of Homo sapiens BAC clone RP11-567N20  
JOURNAL Unpublished (2001)  
REFERENCE 2 (bases 1 to 136657)  
AUTHORS Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (04-MAY-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
REFERENCE 3 (bases 1 to 136657)  
AUTHORS Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (29-AUG-2001) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
REFERENCE 4 (bases 1 to 136657)  
AUTHORS Waterston,R.  
TITLE Direct Submission  
JOURNAL Submitted (09-JAN-2002) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
REFERENCE 5 (bases 1 to 136657)  
AUTHORS Wilson,R.K.  
TITLE Direct Submission  
JOURNAL Submitted (30-APR-2005) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
COMMENT On Aug 29, 2001 this sequence version replaced gi:15144353.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Web site: WUGSC  
Contact: submissions@watson.wustl.edu  
----- Summary Statistics  
-----  
Center project name: H\_NH0567N20  
-----  
NOTICE:  
This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate

chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

#### SOURCE INFORMATION:

The RPEC11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org> VECTOR: pBACe3.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-249F13, 2000 bp overlap; the clone sequenced to the right is RP11-16D24. Actual start of this clone is at base position 149708 of RP11-249F13; actual end is at base position 136657 of RP11-567N20.

Data from AC011936 was used to finish the clone, AC068600. Polymorphisms have been identified between AC011936 and AC068600.

#### FEATURES

source

Location/Qualifiers  
1..136657  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="2"  
/clone="RP11-567N20"  
/clone\_lib="RFC1-11"

#### ORIGIN

Query Match 53.3%; Score 22.4; DB 5; Length 136657;  
Best Local Similarity 72.5%; Pred. No. 6.4e+02;  
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;  
  
Qy 3 CCACGCTTGGCGAGCCCTTACAAAGACTATCAAGTAAAG 42  
Db 59046 CAAAACTAGACCAAGGCTTTACAAAACCTAAGAGTATAG 59007

#### RESULT 76

AC006732

LOCUS

DEFINITION

9 unordered pieces.

AC006732

VERSION

HTG; HTGS PHASE1.

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

JOURNAL

TITLE

REFERENCE

AUTHORS

JOURNAL

TITLE

COMMENT

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 9 contigs. The true order of the pieces

\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 2042: contig of 2042 bp in length  
2043 2056: gap of unknown length  
2057 5001: contig of 2945 bp in length  
5002 5015: gap of unknown length  
5016 11580: contig of 6565 bp in length  
11581 11594: gap of unknown length  
11595 20569: contig of 8975 bp in length  
20570 20583: gap of unknown length  
20584 42974: contig of 22391 bp in length  
42975 42988: gap of unknown length  
42989 59017: contig of 16029 bp in length  
59018 59031: gap of unknown length  
59032 88258: contig of 29227 bp in length  
88259 88272: gap of unknown length  
88273 117204: contig of 28932 bp in length  
117205 117218: gap of unknown length  
117219 159453: contig of 42235 bp in length.

#### FEATURES

source

Location/Qualifiers  
1..159453  
/organism="Caenorhabditis elegans"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:6239"  
/clone="Y32G9"  
2043..2056  
/estimated\_length=unknown  
5002..5015  
/estimated\_length=unknown  
11581..11594  
/estimated\_length=unknown  
20570..20583  
/estimated\_length=unknown  
42975..42988  
/estimated\_length=unknown  
59018..59031  
/estimated\_length=unknown  
88259..88272  
/estimated\_length=unknown  
117205..117218  
/estimated\_length=unknown

#### ORIGIN

Query Match 53.3%; Score 22.4; DB 12; Length 159453;  
Best Local Similarity 81.2%; Pred. No. 6.8e+02;  
Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGGCGAGCCCTTACAAAGACTA 32

Db 21125 GCTCACCTTGGCCTTAGGCTTAACAAAGACTA 21156

#### RESULT 77

AC025518

LOCUS

DEFINITION

AC025518

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

2 (bases 1 to 176355)

AC025518 176355 bp DNA linear PRI 29-JUN-2002  
Homo sapiens chromosome 17, clone RP11-713H12, complete sequence.

AC025518.8 GI:21629418

HTG.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.

1 (bases 1 to 176355)

Birren, B., Nusbaum, C. and Lander, E.

Homo sapiens chromosome 17, clone RP11-713H12

Unpublished

2 (bases 1 to 176355)

## AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F., Boguslavskiy, L., Boukhgalter, B., Brown, A., Burkett, G., Collins, S., Campoliano, A., Castle, A., Choepe, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., Pisanic, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE  
JOURNAL

Submitted (09-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 176355)

REFERENCE  
AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B., Brown, A., Camarata, J., Campoliano, A., Chang, J., Chazaro, B., Choepe, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., FitzGerald, M., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Straus, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

## Direct Submission

Submitted (15-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
4 (bases 1 to 176355)

REFERENCE  
AUTHORS

Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Chazaro, B., Choepe, Y., Collymore, A., Cooke, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Liu, G., MacLean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

## Direct Submission

Submitted (29-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jun 29, 2002 this sequence version replaced gi:21427869.

TITLE  
JOURNAL

## COMMENT

All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence submissions@genome.wi.mit.edu

----- Project Information

Center project name: L8285

Center clone name: 713\_H\_12

## FEATURES

Source	Location/Qualifiers
1..176355	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosomes="17"
	/map="17"
	/clone="RP11-713H12"
	/clone_lib="RPC1-11 Human Male BAC"
complement(30..88)	/rpt_family="L3"
89..388	/rpt_family="AluY"
	/complement(389..503)
	/rpt_family="L3"
591..640	/rpt_family="L2"
1380..2441	/rpt_family="L2"
	/complement(2781..2876)
	/rpt_family="L2"
3360..3655	/rpt_family="AluJb"
4230..4380	/rpt_family="AluY"
4449..4745	/rpt_family="AluSg"
4931..4987	/rpt_family="TAAA"
5006..5074	/rpt_family="TTTC"
	/complement(5183..5252)
	/rpt_family="HAL1"
5680..5735	/rpt_family="CT-rich"
5915..6212	/rpt_family="AluSx"
6636..6943	/rpt_family="AluSp"
	/complement(7166..7384)
	/rpt_family="LIME1"
	/complement(7385..7732)
	/rpt_family="MER2"
	/complement(7733..8912)
	/rpt_family="LIME1"
8913..8940	/rpt_family="TTTG"
	/complement(8941..9394)
	/rpt_family="LIME1"
9395..9432	/rpt_family="TA"
	/complement(9433..10165)
	/rpt_family="LIME1"
	/complement(10889..10989)
	/rpt_family="MIR"
11195..11492	/rpt_family="AluY"
11755..12059	/rpt_family="AluJb"
12354..12497	/rpt_family="MIR"

```

repeat_region complement(12811..12979)
/rpt_family="MERSA"
repeat_region complement(13572..13634)
/rpt_family="L2"
repeat_region complement(13667..14961)
/rpt_family="L2"
repeat_region complement(15020..15149)
/rpt_family="MLT1H"
repeat_region complement(15837..15949)
/rpt_family="L1MC4"
repeat_region complement(15990..16091)
/rpt_family="MER34"
repeat_region complement(16528..16681)
/rpt_family="L1ME3"
repeat_region 16953..17049
/rpt_family="LTR49"
repeat_region 17060..17202
/rpt_family="TA)n"
repeat_region 17224..17534
/rpt_family="Alu9p"
repeat_region 17609..18017
/rpt_family="LTR56"
repeat_region 18788..18850
/rpt_family="MIR"
repeat_region 18911..18947
/rpt_family="MER39B"
repeat_region 20262..20301
/rpt_family="(CAT)n"
repeat_region complement(20318..20535)
/rpt_family="MIR"
repeat_region complement(20626..20718)
/rpt_family="MIR"
repeat_region 21147..21645
/rpt_family="L2"
repeat_region complement(21797..21897)
/rpt_family="L2"
repeat_region complement(22229..22388)
/rpt_family="Alu9/x"
repeat_region complement(22822..22940)

Query Match 53.3%; Score 22.4; DB 5; Length 176355;
Best Local Similarity 72.5%; Pred.No.7e+02; 11; Indels 0; Gaps 0;
Matches 29; Conservative 0; Mismatches 0;

Qy 3 CCAGCTTGGCGGAGCCCTACAAGACTATGAAGTAAAG 42
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 59526 CCAGGCTGGCTGAGCTCTCAGCAATTCATGCACTAAG 59565

RESULT 78
AC024640/c
LOCUS AC024640
DEFINITION Homo sapiens chromosome 2 clone RP11-759B6 map 2, WORKING DRAFT
SEQUENCE 181914 bp DNA linear HTG 24-AUG-2002
ACCESSION AC024640
VERSION AC024640.3 GI:8076901
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.
REFERENCE 1 (bases 1 to 181914)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
TITLE Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
JOURNAL Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
TITLE Campopiano,A., Castelle,A., Choepel,Y., Colangelo,M., Collins,S.,
JOURNAL Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
REFERENCE Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
* NOTE: This is a 'working draft' sequence. It currently
* consists of 25 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence

```

Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczký,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J., Meneus,B., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission  
Submitted (01-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 181914)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T., Lehoczký,J., Levine,R., Liu,G., McEwan,P., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M., Meldrim,J., Meneus,B., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission  
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On May 25, 2000 this sequence version replaced gi:7239651.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L6847  
Center clone name: 759\_B\_6  
----- Summary Statistics  
Sequencing vector: M13; M7815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 165475 bases at least Q40  
Consensus quality: 173706 bases at least Q30  
Consensus quality: 177163 bases at least Q20  
Insert size: 188000; agarose-fp  
Insert size: 179514; sum-of-contigs  
Quality coverage: 3.8 in Q20 bases; agarose-fp  
Quality coverage: 3.9 in Q20 bases; sum-of-contigs  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 25 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence





Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,  
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,  
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,  
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,  
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

## TITLE

Submitted (25-MAR-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA

## REFERENCE

3 (bases 1 to 186870)

## AUTHORS

Barren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,  
Barnia,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B.,  
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,  
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,  
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,  
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,  
Karatas,A., Kellis,C., Landers,T., Levine,R., Lindblad-Toh,K.,  
Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,  
McCarthy,M., Meldrum,J., Meneus,L., Mihova,T., Mlenga,V.,  
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,  
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,  
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,  
Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,  
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,  
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,  
Zembek,L., Zimmer,A. and Zody,M.

## TITLE

Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA

## COMMENT

On Aug 24, 2002 this sequence version replaced gi:19703298.  
All repeats were identified using RepeatMasker:

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L26130

Center clone name: 855 E 10

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 185544 bases at least Q40

Consensus quality: 185963 bases at least Q30

Consensus quality: 186196 bases at least Q20

Insert size: 188000; agarose-fp

Insert size: 186570; sum-of-contigs

Quality coverage: 17.0 in Q20 bases; agarose-fp

Quality coverage: 17.1 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 4 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved

\* 1 16661: contig of 16661 bp in length

\* 16662 16761: gap of 100 bp

\* 16762 42342: contig of 25581 bp in length

\* 42343 42442: gap of 100 bp

\* 42443 114057: contig of 71615 bp in length

\* 114058 114157: gap of 100 bp

\* 114158 186870: contig of 72713 bp in length.

## FEATURES

## source

Location/Qualifiers  
1. 186870  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="17"  
/map="17"  
/clone="RP11-855B10"  
/clone\_lib="RPC1-11 Human Male BAC"

## misc\_feature

1. 16661  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:left  
16662..16761  
/estimated\_length=100  
16762..42342  
/note="assembly\_fragment"  
42343..42442  
/estimated\_length=100  
42443..114057  
/note="assembly\_fragment"  
114058..114157  
/estimated\_length=100  
114158..186870  
/note="assembly\_fragment"

## ORIGIN

Query Match 53.3%; Score 22.4; DB 12; Length 186870;  
Best Local Similarity 72.5%; Pred. No. 7.2e+02;  
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 3 CCAGCTTGGCGAGCCCTTACAAAGACTATGAGTAAAG 42

Db 87951 CCAGCGCTGGTGAGCTCTCAGCAATTCATGCGAGTAAAG 87912

RESULT 80

AX400400

LOCUS

AX400400

DEFINITION

Sequence 76 from Patent WO0210453.

ACCESSION

AX400400

VERSION

AX400400.1

GI:21336580

KEYWORDS

Rattus norvegicus (Norway rat)

SOURCE

Rattus norvegicus

ORGANISM

Rattus norvegicus

REFERENCE

1

AUTHORS

Mendrick,D., Porter,M.W., Johnson,K.R., Castle,A.L. and

Elashoff,M.R.

TITLE

Molecular toxicology modeling

JOURNAL

Patent: WO 0210453-A 76 07-FEB-2002;

Gene Logic, Inc. (US)

FEATURES

Location/Qualifiers

source

1. 584

/organism="Rattus norvegicus"

/mol\_type="unassigned DNA"

/db\_xref="taxon:10116"

/note="EMBL/GenBank Accession No. AA818700"

ORIGIN

Query Match 52.9%; Score 22.2; DB 2; Length 584;

Best Local Similarity 88.9%; Pred. No. 1.2e+02;

Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 15 GAGCCCTTACAAAGACTATGAGTAAA 41

Db 373 GAGCCCTTACAAAGACTATGAGTAAA 399

RESULT 81

AC104541

```

LOCUS      AC104541                      69203 bp      DNA      linear      HTG 13-DEC-2001
DEFINITION Mus musculus clone RP23-142P5, LOW-PASS SEQUENCE SAMPLING.
ACCESSION  AC104541
VERSION    AC104541.1  GI:17646941
KEYWORDS   HTG; HTGS PHASE0.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidae; Muridae; Murinae; Mus.
1 (bases 1 to 69203)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Mus musculus, clone RP23-142P5
Unpublished
2 (bases 1 to 69203)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,I., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepe,Y., Colangelo,M., Collins,S., Collamore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRoque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Mihov,I., Mienga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Triglio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (13-DEC-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genom Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L19844
Center clone name: 142_P_5
-----
* NOTE: This record contains 81 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1 764: contig of 764 bp in length
* 765 864: gap of 100 bp
* 865 1620: contig of 756 bp in length
* 1621 1720: gap of 100 bp
* 1721 2493: contig of 773 bp in length
* 2494 2593: gap of 100 bp
* 2594 3317: contig of 724 bp in length
* 3318 3417: gap of 100 bp
* 3418 4160: contig of 743 bp in length
* 4161 4260: gap of 100 bp
*
* 4261 5031: contig of 771 bp in length
* 5032 5131: gap of 100 bp
* 5132 5900: contig of 769 bp in length
* 5901 6000: gap of 100 bp
* 6001 6734: contig of 734 bp in length
* 6735 7598: contig of 764 bp in length
* 7599 7699: gap of 100 bp
* 7699 8457: contig of 759 bp in length
* 8458 8557: gap of 100 bp
* 8558 9286: contig of 729 bp in length
* 9287 9386: gap of 100 bp
* 9387 10132: contig of 746 bp in length
* 10133 10232: gap of 100 bp
* 10233 10989: contig of 757 bp in length
* 10990 11090: gap of 100 bp
* 11090 11970: contig of 781 bp in length
* 11971 12752: contig of 782 bp in length
* 12753 12852: gap of 100 bp
* 12853 13311: contig of 779 bp in length
* 13312 13731: gap of 100 bp
* 13732 14493: contig of 762 bp in length
* 14494 15316: contig of 723 bp in length
* 15317 15416: gap of 100 bp
* 15417 16335: contig of 719 bp in length
* 16336 16235: gap of 100 bp
* 16236 17003: contig of 768 bp in length
* 17004 17103: gap of 100 bp
* 17104 17877: contig of 774 bp in length
* 17878 17977: gap of 100 bp
* 17978 18077: contig of 730 bp in length
* 18078 18807: gap of 100 bp
* 18808 19548: contig of 741 bp in length
* 19549 19648: gap of 100 bp
* 19649 20362: contig of 714 bp in length
* 20363 20462: gap of 100 bp
* 20463 21221: contig of 759 bp in length
* 21222 21321: gap of 100 bp
* 21322 22056: contig of 735 bp in length
* 22057 22156: gap of 100 bp
* 22157 22926: contig of 770 bp in length
* 22927 23026: gap of 100 bp
* 23027 23793: contig of 767 bp in length
* 23794 23893: gap of 100 bp
* 23894 24634: contig of 741 bp in length
* 24635 24734: gap of 100 bp
* 24735 25495: contig of 761 bp in length
* 25496 25595: gap of 100 bp
* 25596 26356: contig of 761 bp in length
* 26357 26456: gap of 100 bp
* 26457 27197: contig of 741 bp in length
* 27198 27297: gap of 100 bp
* 27298 28039: contig of 742 bp in length
* 28040 28139: gap of 100 bp
* 28140 28887: contig of 748 bp in length
* 28888 28987: gap of 100 bp
* 28988 29760: contig of 773 bp in length
* 29761 29860: gap of 100 bp
* 29861 30636: contig of 776 bp in length
* 30637 30736: gap of 100 bp
* 30737 31500: contig of 764 bp in length
* 31501 31600: gap of 100 bp
* 31601 32358: contig of 758 bp in length
* 32359 32458: gap of 100 bp
* 32459 33229: contig of 771 bp in length
* 33230 33329: gap of 100 bp
* 33330 34052: contig of 723 bp in length
* 34053 34152: gap of 100 bp
* 34153 34911: contig of 759 bp in length
* 34912 35011: gap of 100 bp
* 35012 35744: contig of 733 bp in length

```

\* 35745 35844: gap of 100 bp  
\* 35845 3615: contig of 771 bp in length  
\* 3616 36715: gap of 100 bp  
\* 36716 37453: contig of 738 bp in length  
\* 37454 37553: gap of 100 bp  
\* 37554 38275: contig of 722 bp in length  
\* 38276 38375: gap of 100 bp  
\* 38376 39097: contig of 722 bp in length  
\* 39098 39197: gap of 100 bp  
\* 39198 39943: contig of 746 bp in length  
\* 39944 40043: gap of 100 bp  
\* 40044 40816: contig of 773 bp in length  
\* 40817 40916: gap of 100 bp  
\* 40917 41681: contig of 765 bp in length  
\* 41682 41781: gap of 100 bp  
\* 41782 42545: contig of 764 bp in length  
\* 42546 42645: gap of 100 bp  
\* 42646 43395: contig of 750 bp in length  
\* 43396 43495: gap of 100 bp  
\* 43496 44248: contig of 753 bp in length  
\* 44249 44348: gap of 100 bp  
\* 44349 45112: contig of 764 bp in length  
\* 45113 45212: gap of 100 bp  
\* 45213 45983: contig of 771 bp in length  
\* 45984 46083: gap of 100 bp  
\* 46084 46831: contig of 748 bp in length  
\* 46832 46931: gap of 100 bp  
\* 46933 47714: contig of 783 bp in length  
\* 47715 47814: gap of 100 bp  
\* 47815 48580: contig of 766 bp in length  
\* 48581 48680: gap of 100 bp  
\* 48681 49462: contig of 782 bp in length  
\* 49463 49562: gap of 100 bp  
\* 49563 50338: contig of 776 bp in length  
\* 50339 50439: gap of 100 bp  
\* 50439 51159: contig of 721 bp in length  
\* 51160 51259: gap of 100 bp  
\* 51260 52017: contig of 758 bp in length  
\* 52018 52117: gap of 100 bp  
\* 52118 52883: contig of 766 bp in length  
\* 52884 52983: gap of 100 bp  
\* 52984 53752: contig of 769 bp in length  
\* 53753 53852: gap of 100 bp  
\* 53853 54593: contig of 741 bp in length  
\* 54594 54693: gap of 100 bp  
\* 54694 55434: contig of 741 bp in length  
\* 55435 55534: gap of 100 bp  
\* 55535 56261: contig of 727 bp in length  
\* 56262 56361: gap of 100 bp  
\* 56362 57100: contig of 739 bp in length  
\* 57101 57200: gap of 100 bp  
\* 57201 57969: contig of 769 bp in length  
\* 57970 58069: gap of 100 bp  
\* 58070 58840: contig of 771 bp in length

Query Match 52.9%; Score 22.2; DB 12; Length 69203;  
Best Local Similarity 77.1%; Pred. No. 6.3e-02;  
Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 7 GCTTGGCGAGCCCTTACAAAGACTATGAAGTAAA 41  
|||||  
Db 53656 GCTTGGCAAGCTCTTAGAAGATACATAAGTAAA 53690

RESULT 82

AC011600\_3

WPCOMMENT

Sequence split into 4 fragments LOCUS AC011600 Accession AC011600  
Fragment Name Begin End  
AC011600\_0 1 110000  
AC011600\_1 100001 210000  
AC011600\_2 200001 310000  
AC011600\_3 300001 373878

Continuation (4 of 4) of AC011600 from base 300001 (AC011600 Homo sapiens chromosome 12 c  
Query Match 52.9%; Score 22.2; DB 12; Length 73878;  
Best Local Similarity 77.1%; Pred. No. 6.4e-02;  
Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
Qy 7 GCTTGGCGAGCCCTTACAAAGACTATGAAGTAAA 41  
|||||  
Db 4908 GCCTTGACAGTTATTACAAAGATATGAAGTAAA 4942

RESULT 83

AC011600\_2

WPCOMMENT

Sequence split into 4 fragments LOCUS AC011600 Accession AC011600  
Fragment Name Begin End  
AC011600\_0 1 110000  
AC011600\_1 100001 210000  
AC011600\_2 200001 310000  
AC011600\_3 300001 373878

Continuation (3 of 4) of AC011600 from base 200001 (AC011600 Homo sapiens chromosome 12 c

Query Match 52.9%; Score 22.2; DB 12; Length 110000;  
Best Local Similarity 77.1%; Pred. No. 7.4e-02;  
Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 7 GCTTGGCGAGCCCTTACAAAGACTATGAAGTAAA 41  
|||||  
Db 104908 GCCTTGACAGTTATTACAAAGATATGAAGTAAA 104942

RESULT 84

AC079597/c

LOCUS

AC079597 156766 bp DNA linear PRI 25-SEP-2002  
DEFINITION Homo sapiens 12q BAC RP11-324H9 (Roswell Park Cancer Institute  
Human BAC Library) complete sequence.

ACCESSION AC079597.13 GI:13129401

VERSION HTG.

KEYWORDS Homo sapiens (human)

SOURCE Homo sapiens

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.

1 (bases 1 to 156766)

REFERENCE  
AUTHORS  
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,  
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,  
Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,  
Bowie,S., Brieve,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,  
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,  
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,  
Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D.,  
Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,  
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,  
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,  
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,  
Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D.,  
Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J.,  
Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W.,  
Gunaratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K.,  
Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O.,  
Hodgson,A., Hognes,M., Holloway,C., Hollins,B., Honsi,F.,  
Howard,S., Huber,J., Hulyk,S., Hume,J., Ioshikhes,I., Jackson,L.E.,  
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,  
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,  
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C.,  
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,  
Louisseg,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,  
Ma,J., Maheshwari,M., Mapua,P., Marondel,I., Martin,R.,  
Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P.,  
Meador,M., Mei,G., Merscher,S., Metzker,M., Miller,A., Miner,G.,  
Miner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M.,  
Morris,S., Moser,M., Neal,D., Nelson,D., Newton,J., Newton,N.,



```

/function="clone overlap"
complement(1..44)
/rpt_family="THE1C"
complement(45..399)
/rpt_family="THE1C-internal"
971..1076
/rpt_family="MIR"
2309..2364
/rpt_family=" (CATATA)n"
2365..2444
/rpt_family="MER47"
2493..2559
/rpt_family="MER47"
2569..2796

Query Match      52.9%; Score 22.2; DB 5; Length 156766;
Best Local Similarity 77.1%; Pred. No. 8.4e+02;
Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 7 GCTTGGCGAGCCCTTACAAAGACTATGAAGTAA 41
Db 70179 GCCTTGCAAGTTATTACAAAGAATATGAAGTAA 70145

RESULT 85
AC128161/c
LOCUS AC128161 205274 bp DNA linear HTG 22-SEP-2002
DEFINITION Rattus norvegicus clone CH230-58G18, *** SEQUENCING IN PROGRESS
***, 5 unordered pieces.
ACCESSION AC128161 GI:23196340
VERSION AC128161.2 HTGS PHASE3; HTGS DRAFT; HTGS_ENRICHED.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
1 (bases 1 to 205274)
Muzny,D,Marie., Metzker,M,Lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alebrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biewald,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,K., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,
Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Lu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Maheshwari,M., Mahindartine,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,G., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwackeleleh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.,
Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,

```

Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,  
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,  
Sanders,W., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,  
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,  
Sneed,A., Sodergren,E., Song,X.-Z., Sorrelle,R., Sosa,J., Taylor,C.,  
Teimle,M., Strong,R., Sutton,A., Svatek,A., Tabot,P., Taylor,C.,  
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Uemami,K.,  
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,  
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,  
Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,  
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,  
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von  
Wiederhauser,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,  
Weinstock,G. and Gibbs,R.A.

Unpublished  
Direct Submission  
2 (bases 1 to 205274)  
Worley,K.C.  
Direct Submission  
Submitted (19-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 205274)  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (22-SEP-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Sep 19, 2002 this sequence version replaced gi:21908761.  
The sequence in this assembly is a combination of BAC based reads  
and whole genome shotgun sequencing reads assembled using Atlas  
(http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the  
sequence may extend beyond the ends of the clone and there may be  
contigs that consist entirely of whole genome shotgun sequence  
reads. Both end sequences and whole genome shotgun sequence only  
contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: http://www.hgsc.bcm.tmc.edu/  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GRAV  
Center clone name: CH230-58G18  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 185782 bases at least Q40  
Consensus quality: 188579 bases at least Q30  
Consensus quality: 190025 bases at least Q20  
Estimated insert size: 210634; sum-of-contigs estimation  
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 5 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
1 13135: contig of 13135 bp in length  
\* 13136 13235: gap of unknown length  
\* 13236 191447: contig of 178212 bp in length  
\* 191448 191547: gap of unknown length  
\* 191548 192710: contig of 1163 bp in length  
\* 192711 192810: gap of unknown length  
\* 192811 197250: contig of 4440 bp in length  
\* 197251 197350: gap of unknown length  
\* 197351 205274: contig of 7924 bp in length.  
FEATURES  
Location/Qualifiers  
1..205274  
source

```

/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-58G18"
13136..13235
/estimated_length=unknown
complement(152564..178037)
/note="clone_boundary
clone end:T7
site:ECORI
end_sequence:BH302924"
191448..191547
/estimated_length=unknown
192711..192810
/estimated_length=unknown
197251..197350
/estimated_length=unknown

gap
misc_feature

gap
gap
gap

ORIGIN
Query Match 52.9%; Score 22.2; DB 12; Length 205274;
Best Local Similarity 88.9%; Pred. No. 9.2e+02;
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 15 GAGCCCTTACAAAGACTATGAAGTAAA 41
Db 119277 GAGCCCTTACAAAGACTAAGAACCAA 119251
|||||

```

```

restriction digest.
SOURCE INFORMATION:

```

This sequence is the entire insert of the clone.

```

FEATURES
    source
        1..216886
            /organism="Gallus gallus"
            /mol_type="genomic DNA"
            /db_xref="taxon:9031"
            /chromosome="ul"
            /clone="CH261-11F14"
            /clone_lib="CHORI261"

```

## ORIGIN

```

Query Match 52.9%; Score 22.2; DB 11; Length 216886;
Best Local Similarity 77.1%; Pred. No. 9.4e+02;
Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```

```

QY 8 CTGCGCGAGCCCTTACAAAGACTATGAAGTAAAG 42
Db 91252 CTGCTGTGAGCACTTCCAAAGGATATGAAATAAAG 91218
|||||

```

## RESULT 87

AC096068/c

```

LOCUS
DEFINITION
    Rattus norvegicus clone CH230-20F15, *** SEQUENCING IN PROGRESS
    ***, 4 unordered pieces.

```

```

ACCESSION
    AC096068.6 GI:24817916
KEYWORDS
    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE
    Rattus norvegicus (Norway rat)

```

## ORGANISM

Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridea; Muridae; Murinae; Rattus.

## REFERENCE

## AUTHORS

Muzny, D., Marie, J., Metzker, M., Lee, J., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A., Chacko, J., Chavez, D., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Huiyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuewa, L., Louiseged, H., Lozada, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Milosavljevic, A., Miner, G., McNeill, T. Z., Meenen, E., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Naik, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwackeleme, O., Okwunonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,

## NOTICE:

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by

Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,  
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
 Rivers, C., Rockey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,  
 Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,  
 Shetty, J., Shvartsbeyn, A., Sigson, I., Sitter, C.D., Snajls, D.,  
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,  
 Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,  
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,  
 Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,  
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,  
 Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K.,  
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von  
 Niederhausern, A., Weiss, R., Smith, D.R., Hoit, R.A., Smith, H.O.,  
 Weinstock, G. and Gibbs, R.A.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 226291)  
 Worley, K.C.  
 Direct Submission  
 Submitted (17-SEP-2001) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 226291)  
 Rat Genome Sequencing Consortium.  
 Direct Submission  
 Submitted (09-NOV-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Nov 9, 2002 this sequence version replaced gi:22772219.  
 The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described  
 in the feature table below represents a scaffold in the Atlas  
 assembly (a 'contig-scaffold'). Within each contig-scaffold,  
 individual sequence contigs are ordered and oriented, and separated  
 by sized gaps filled with Ns to the estimated size. The sequence  
 may extend beyond the ends of the clone and there may be sequence  
 contigs within a contig-scaffold that consist entirely of whole  
 genome shotgun sequence reads. Both end sequences and whole genome  
 shotgun sequence only contigs will be indicated in the feature  
 table.

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

COMMENT

----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GSI1  
 Center clone name: CH230-20F15  
 ----- Summary Statistics  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 183145 bases at least Q40  
 Consensus quality: 187692 bases at least Q30  
 Consensus quality: 189335 bases at least Q20  
 Estimated insert size: 184258; sum-of-contigs estimation  
 Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

-----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 \* NOTE: this is a 'working draft' sequence. It currently  
 \* consists of 4 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence.  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 220799: contig of 220799 bp in length  
 \* 220800 220899: gap of unknown length  
 \* 220900 221990: contig of 1091 bp in length  
 \* 221991 222090: gap of unknown length  
 \* 222091 223397: contig of 1307 bp in length

\* 223398 223497: gap of unknown length  
 \* 223498 226291: contig of 2794 bp in length.  
 FEATURES  
 source

1. 226291  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-20F15"

misc\_feature

1. 835  
 /note="clone\_boundary"  
 clone end: 77  
 site: EcoRI

misc\_feature

end sequence: BH319833"  
 174564..175853

misc\_feature

/note="wgs contig"  
 217818..220799

gap

/note="wgs contig"  
 220800..220899

gap

/estimated\_length=unknown  
 221991..222090

gap

/estimated\_length=unknown  
 223398..223497

gap

/estimated\_length=unknown

ORIGIN

Query Match 52.9%; Score 22.2; DB 12; Length 226291;  
 Best Local Similarity 88.9%; Pred. No. 9.5e+02;  
 Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 15 GAGCCCTTACAAAGACTATGAAGTAAA 41  
 |||||

Db 52562 GAGCCCTTACAAAGACTAAGAACCAA 52536  
 |||||

RESULT 88

AC126713 AC126713 247515 bp DNA linear HTG 09-MAY-2003  
 LOCUS Rattus norvegicus clone CH230-5L14, WORKING DRAFT SEQUENCE, 6  
 DEFINITION  
 unorderd pieces.

ACCESSION AC126713.4 GI:30466573

VERSION HTG: HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.

KEYWORDS Rattus norvegicus (Norway rat)

SOURCE

ORGANISM  
 Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muroidea; Muridae; Murinae; Rattus.

REFERENCE

AUTHORS

1 (bases 1 to 247515)  
 Murny, D., Marie, Metzker, M., Lee, Abranzon, S., Adams, C., Alder, J.,  
 Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,  
 Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,  
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,  
 Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
 Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,  
 Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,  
 Chacko, J., Chavez, D., Chen, R., Chen, Y., Chen, Z., Chu, J.,  
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
 Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,  
 Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,  
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,  
 Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,  
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
 Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
 Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guera, W.,  
 Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,  
 Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
 Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hoques, M.,  
 Hollins, B., Howells, S., Huliy, S., Hume, J., Idlebird, D., Jackson, A.,  
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,  
 Karpachy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,  
 Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
 Lorensuhewa, L., Louised, H., Lozada, R.J., Lu, X., Ma, J.,



Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokelimeh, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartabeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steinkle, M., Strong, R., Sutton, A., Svatek, A., Tabor, F., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umani, K., Valdes, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

# TITLE JOURNAL REFERENCE AUTHORS TITLE JOURNAL

Direct Submission  
 Unpublished  
 2 (bases 1 to 247515)  
 Worley, K.C.

Submitted (09-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

# REFERENCE AUTHORS TITLE JOURNAL

Rat Genome Sequencing Consortium.  
 Direct Submission

Submitted (09-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

# COMMENT

On May 9, 2003 this sequence version replaced gi:24941703. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information

Center project name: GBK0  
 Center clone name: CH230-5L14  
 ----- Summary Statistics

Assembly program: Atlas;  
 Consensus quality: 228033 bases at least Q40  
 Consensus quality: 230259 bases at least Q30  
 Consensus quality: 231440 bases at least Q20  
 Estimated insert size: 237542; sum-of-contigs estimation  
 Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

-----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 6 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1 236249: contig of 236249 bp in length  
 \* 236250 236349: gap of unknown length  
 \* 236350 237425: contig of 1076 bp in length  
 \* 237426 237525: gap of unknown length  
 \* 237526 239390: contig of 1865 bp in length  
 \* 239391 239491: gap of unknown length  
 \* 239491 240713: contig of 1223 bp in length  
 \* 240714 240813: gap of unknown length  
 \* 240814 242869: contig of 2056 bp in length  
 \* 242870 242969: gap of unknown length  
 \* 242970 247515: contig of 4546 bp in length.

# FEATURES

source  
 1..247515  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-5L14"

# misc\_feature

1..1818  
 /note="wgs\_end\_extension  
 clone\_end:T7"

# misc\_feature

12486..14214  
 /note="wgs\_end\_extension  
 clone\_end:T7"

# misc\_feature

15239..15971  
 /note="clone\_boundary  
 clone\_end:T7"

# misc\_feature

end\_sequence:BH362758"  
 complement(233573..234468)  
 /note="clone\_boundary  
 clone\_end:Sp6  
 site:EcoRI

end\_sequence:BH362760"  
 236250..236349  
 /estimated\_length=unknown

# gap

237426..237525  
 /estimated\_length=unknown

# gap

239391..239490  
 /estimated\_length=unknown

# gap

240714..240813  
 /estimated\_length=unknown

# gap

242870..242969  
 /estimated\_length=unknown

# ORIGIN

Query Match 52.9%; Score 22.2; DB 12; Length 247515;  
 Best Local Similarity 77.1%; Pred.No.9.8e+02;  
 Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 2 CCCAGCGTTGGCCGAGCCCTTACAAAGACTATGAA 36

Db 95194 CACAGGAGTGGCAGACCTACAGAGACTATGAA 95228

# RESULT 89

AC122818

# LOCUS

AC122818 167735 bp DNA linear ROD 08-NOV-2003

# DEFINITION

Mus musculus BAC clone RP23-243016 from 17, complete sequence.

# ACCESSION

AC122818

# VERSION

AC122818.4 GI:26801343

# KEYWORDS

HTG.

# SOURCE

Mus musculus (house mouse)

# ORGANISM

Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 167735)

Trani, L., Kozlowski, A. and Schatzkammer, K.

The sequence of Mus musculus BAC clone RP23-243016



```

repeat_region 11811..11899
/rpt family="B2"
repeat_region 11907..12036
/rpt family="Alu"
repeat_region 12118..12302
/rpt family="B2"
repeat_region 12319..12428
/rpt family="Alu"
repeat_region 12817..13011
/rpt family="B2"
repeat_region 13559..13891
/rpt family="ERVK"
repeat_region 13996..14381
/rpt family="RMER198"
repeat_region 15092..16198
/rpt family="ERVK"
repeat_region 16595..16887
/rpt family="ERVK"
repeat_region 17012..17191
/rpt family="B2"
repeat_region 17228..17404
/rpt family="L1"
repeat_region 17407..17466
/rpt family="B4"
repeat_region 18377..18803
/rpt family="ERVK"
repeat_region 18802..18885
/rpt family="ERVK"
repeat_region 18884..19056
/rpt family="ERVK"
repeat_region 19127..19448
/rpt family="ERVK"
repeat_region 19811..19900
/rpt family="B4"
repeat_region 19970..20101
/rpt family="Alu"
repeat_region 20294..20437
/rpt family="Alu"
repeat_region 20532..20721
/rpt family="B2"
repeat_region 20960..21390

Query Match 52.4%; Score 22; DB 6; Length 167735;
Best Local Similarity 83.3%; Pred.No. 1.1e+03;
Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 13 CCGAGCCCTTACAAAGACTATGAGGTAAG 42
Db 160901 CCAAGCCAGTACAAAGACAATGAGGTAAG 160930

RESULT 90
AC121302/c
LOCUS AC121302 190256 bp DNA linear ROD 28-JAN-2004
DEFINITION Mus musculus chromosome 13, clone RP23-163B12, complete sequence.
ACCESSION AC121302
VERSION AC121302.7 GI:41351614
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.
1 (bases 1 to 190256)
Birren,B., Nusbaum,C. and Lander,E.
TITLE Mus musculus chromosome 13, clone RP23-163B12
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 190256)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,
Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Collangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Jones,C.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
McDonald,P., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,B., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Submitted (27-NOV-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 190256)
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
McDonald,P., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,B., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (16-MAY-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 190256)
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
McDonald,P., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,B., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (28-JAN-2004) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jan 28, 2004 this sequence version replaced gi:38564369.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center

```

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR  
 Web site: <http://www-seq.wi.mit.edu>  
 Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
 ----- Project Information  
 Center project name: L24214  
 Center clone name: 163\_B12

## FEATURES

```

source
1. .190256
  /organism="Mus musculus"
  /mol_type="genomic DNA"
  /db_xref="taxon:10090"
  /chromosomes="13"
  /map="13"
  /clone="RP23-163B12"
  /clone_lib="RPCI-23 Female Mouse BAC"
1. .573
  /note="wgs end extension"
  clone end:SP6
  complement(103. .872)
  /rpt_family="HAL1"
574. .579
  /note="clone boundary"
  clone end:SP6
  site:ECORI
  /rpt_family="B3A"
2218. .2500
  /rpt_family="RLTR9C"
2568. .2653
  /rpt_family="(TA)n"
2877. .2921
  /rpt_family="AT_rich"
4036. .4088
  /rpt_family="(TG)n"
9881. .10185
  /rpt_family="LIMA4A"
  complement(10613. .10714)
  /rpt_family="ETNERV"
  complement(13182. .13371)
  /rpt_family="B2_Mml1"
  complement(13888. .14073)
  /rpt_family="B4A"
14306. .14327
  /rpt_family="AT_rich"
  complement(14764. .14888)
  /rpt_family="MTE"
  complement(14922. .15099)
  /rpt_family="MTE"
15594. .15790
  /rpt_family="B3"
  complement(15825. .15989)
  /rpt_family="Lx9"
15987. .16183
  /rpt_family="Lx9"
16195. .16275
  /rpt_family="(TA)n"
16330. .16358
  /rpt_family="AT_rich"
  complement(18164. .18249)
  /rpt_family="RMER15"
  complement(18448. .18946)
  /rpt_family="Lx8"
20301. .20478
  /rpt_family="(TG)n"
20501. .20545
  /rpt_family="(TCTG)n"
20748. .20778
  /rpt_family="(A)n"
  complement(20844. .21038)
  /rpt_family="B4A"
21436. .21499

```

```

repeat_region
  /rpt_family="MIR"
21924. .21965
  /rpt_family="(T)n"
22031. .22052
  /rpt_family="AT_rich"
22799. .23516
  /rpt_family="L1_MM"
24908. .24937
  /rpt_family="AT_rich"
25101. .25224
  /rpt_family="RSINE1"
25452. .25499
  /rpt_family="AT_rich"
25600. .25658
  /rpt_family="(TA)n"
25840. .25880
  /rpt_family="AT_rich"
25980. .26081
  /rpt_family="A-rich"
26305. .26337
  /rpt_family="(TTCA)n"
26344. .26384
  /rpt_family="(TAAAA)n"
26424. .26468
  /rpt_family="B1_MM"
  complement(26469. .26703)
  /rpt_family="Lx6"
27377. .27424
  /rpt_family="(TAGA)n"
  complement(28896. .29877)
  /rpt_family="Lx2"
31058. .31242
  /rpt_family="(GAAA)n"
31334. .31462

Query Match      52.4%; Score 22; DB 6; Length 190256;
Best Local Similarity 73.7%; Pred.No. 1.1e+03;
Matches 28; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY      3 CCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAA 40
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      80046 CCACGCTTGGACAGACCCCTGAAAAACTATGATGAAA 80009

RESULT 91
AC122489/c      191710 bp DNA linear ROD 13-NOV-2003
LOCUS      Mus musculus BAC clone RP24-386H10 from chromosome 13, complete
DEFINITION      sequence.
ACCESSION      AC122489
VERSION      AC122489.3 GI:28316670
KEYWORDS      HTG.
SOURCE      Mus musculus (house mouse)
ORGANISM      Mus musculus
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
      Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE
AUTHORS      Dignan,G. and Meyer,R.
TITLE      The sequence of Mus musculus BAC clone RP24-386H10
JOURNAL      Unpublished (2001)
AUTHORS      Wilson,R.
TITLE      Sequencing of Mus musculus
JOURNAL      Unpublished (2001)
AUTHORS      McPherson,J.D. and Waterston,R.H.
TITLE      Direct Submission
JOURNAL      Submitted (01-JUN-2002) Genome Sequencing Center, 4444 Forest Park
      Parkway, St. Louis, MO 63108, USA
REFERENCE
AUTHORS      McPherson,J.D. and Waterston,R.H.
TITLE      Direct Submission

```

JOURNAL  
Submitted (12-FEB-2003) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA  
5 (bases 1 to 191710)  
McPherson,J.D. and Waterston,R.H.  
Direct Submission  
Submitted (28-FEB-2003) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA  
6 (bases 1 to 191710)  
Wilson,R.  
Direct Submission  
Submitted (13-NOV-2003) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Feb 12, 2003 this sequence version replaced gi:22091363.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu>  
Contact: [submissions@watson.wustl.edu](mailto:submissions@watson.wustl.edu)  
----- Summary Statistics  
Center project name: M\_BB0386H10  
-----

NOTICE: This sequence may not represent the entire insert of this  
clone. It may be shorter because we only sequence overlapping  
clone sections once, or longer because we provide a small overlap  
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate  
chemistry, or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by sequence  
from more than one subclone; and the assembly was confirmed by  
restriction digest.

MAPPING INFORMATION:  
Mapping information for this clone was provided by Dr. Wes Warren,  
Department of Genetics, Washington University, St. Louis MO. For  
additional information about the map position of this sequence, see  
<http://genome.wustl.edu>

SOURCE INFORMATION:  
The RPCI-24 BAC Library has been constructed by Pieter de Jong and  
coworkers (<http://www.chori.org>) from male C57Bl/6J mouse spleen  
and/or brain genomic DNA. The clone and detailed information can be  
obtained from Pieter de Jong and coworkers at <http://www.chori.org>

NEIGHBORING SEQUENCE INFORMATION:  
This sequence is the entire insert of the clone.

FEATURES  
source  
1..191710  
Location/Qualifiers  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="13"  
/map="13"  
/clone="RP24-386H10"  
/clone\_lib="RPCI-24"  
14..308  
repeat\_region  
14..308  
/rpt\_family="MaLR"  
312..393  
repeat\_region  
312..393  
/rpt\_family="RME198"  
722..837  
repeat\_region  
722..837  
/rpt\_family="Alu"  
839..1094  
repeat\_region  
839..1094  
/rpt\_family="L1"  
1355..1438  
repeat\_region  
1355..1438  
/rpt\_family="L1"  
1531..1712  
repeat\_region  
1531..1712  
/rpt\_family="L1"  
2281..2409  
repeat\_region  
2281..2409  
/rpt\_family="B4"  
3565..3639  
repeat\_region  
3565..3639  
/rpt\_family="L1"

repeat\_region  
5003..5217  
/rpt\_family="ID"  
repeat\_region  
6484..6554  
/rpt\_family="B2"  
repeat\_region  
7011..7196  
/rpt\_family="ID"  
repeat\_region  
7217..10174  
/rpt\_family="L1"  
repeat\_region  
10172..10275  
/rpt\_family="L1"  
repeat\_region  
10285..10402  
/rpt\_family="L1"  
repeat\_region  
10406..10661  
/rpt\_family="L1"  
repeat\_region  
10661..10743  
/rpt\_family="L1"  
repeat\_region  
11004..12487  
/rpt\_family="L1"  
repeat\_region  
12492..12769  
/rpt\_family="L1"  
repeat\_region  
12770..13270  
/rpt\_family="ERVK"  
repeat\_region  
13276..13525  
/rpt\_family="ERVK"  
repeat\_region  
13635..13936  
/rpt\_family="ERVK"  
repeat\_region  
15387..16293  
/rpt\_family="ERVK"  
repeat\_region  
17018..17297  
/rpt\_family="ERVK"  
repeat\_region  
17335..17438  
/rpt\_family="ERVK"  
repeat\_region  
18297..18485  
/rpt\_family="ERVK"  
repeat\_region  
18488..18989  
/rpt\_family="ERVK"  
repeat\_region  
18990..19496  
/rpt\_family="L1"  
repeat\_region  
20092..20336  
/rpt\_family="L1"  
repeat\_region  
20635..20737  
/rpt\_family="B4"  
repeat\_region  
22375..22975  
/rpt\_family="L1"  
repeat\_region  
24214..24353  
/rpt\_family="Alu"  
repeat\_region  
25054..25209  
/rpt\_family="ERVK"  
repeat\_region  
26586..27654  
/rpt\_family="L1"  
repeat\_region  
30689..31458  
/rpt\_family="L1"  
repeat\_region  
32580..32749  
/rpt\_family="B2"  
repeat\_region  
32804..33086  
/rpt\_family="ERVK"  
repeat\_region  
37498..37706  
/rpt\_family="MaLR"  
repeat\_region  
40488..40687  
/rpt\_family="L1"  
repeat\_region  
41185..41319  
/rpt\_family="ERVK"  
repeat\_region  
43768..43957  
/rpt\_family="B2"  
repeat\_region  
44474..44659  
/rpt\_family="B4"  
repeat\_region  
45350..45685  
/rpt\_family="MaLR"  
repeat\_region  
46180..46376  
/rpt\_family="B2"  
repeat\_region  
46411..46575  
/rpt\_family="L1"

```

repeat_region      46573..46825
/rpt_family="L1"
repeat_region      49034..49594
/rpt_family="L1"
repeat_region      50897..50945
/rpt_family="ERV1"
repeat_region      51121..51243
/rpt_family="B4"
repeat_region      51430..51624
/rpt_family="B4"
repeat_region      53385..54102
/rpt_family="L1"
repeat_region      55687..55810
/rpt_family="B4"
repeat_region      56725..56852
/rpt_family="L1"
repeat_region      56935..57004
/rpt_family="L1"
repeat_region      57010..57054
/rpt_family="Alu"
repeat_region      57058..57291
/rpt_family="L1"
repeat_region      59494..60476
/rpt_family="L1"
repeat_region      60475..60534
/rpt_family="L1"
repeat_region      61552..61785

Query Match      52.4%; Score 22; DB 6; Length 191710;
Best Local Similarity 73.7%; Pred. No. 1.1e+03;
Matches 28; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 3 CCACGCTTGGCGACGCCCTTACAAAGACTATGAATAA 40
||||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 110632 CCACGCTTGGACAGACCCCTGAAAACTATGATGAAA 110595

RESULT 92
AC084383/c
LOCUS AC084383 203589 bp DNA linear ROD 31-OCT-2003
DEFINITION Mus musculus BAC clone RP23-10B20 from 17, complete sequence.
ACCESSION AC084383
VERSION AC084383.1 GI:15778815
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Nguyen,C., Abbott,A., Elliott,G. and Hawkins,M.
TITLE The sequence of Mus musculus BAC clone RP23-10B20
JOURNAL Unpublished (2001)
REFERENCE
AUTHORS Wilson,R.
TITLE Sequencing of Mus musculus
JOURNAL Unpublished (2001)
REFERENCE
AUTHORS McPherson,J.D.
TITLE Direct Submission
JOURNAL Submitted (26-SEP-2001) Genetics, Genome Sequencing Center, 4444
Forest Park Parkway, St. Louis, Missouri 63108, USA
REFERENCE
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (31-OCT-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT
----- Genome Center

```

Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu>  
Contact: [submissions@watson.wustl.edu](mailto:submissions@watson.wustl.edu)  
----- Summary Statistics  
Center project name: M\_BA0010B20  
-----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclones; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

SOURCE INFORMATION:  
The RP23-10B20 BAC Library has been constructed by Kazutoyo Osegawa and Minako Tateno in the laboratory of Pieter de Jong (<http://www.chori.org>) from female C57BL/6J mouse kidney and/or brain genomic DNA. The clone and detailed information can be obtained from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

NEIGHBORING SEQUENCE INFORMATION:  
This sequence is the entire insert of the clone.

FEATURES	source
repeat_region	1..203589 /organism="Mus musculus" /mol_type="genomic DNA" /db_xref="taxon:10090" /chromosome="17" /map="17" /clone="RP23-10B20" /clone_lib="RPCI-23"
repeat_region	46..184 /rpt_family="B4"
repeat_region	292..1382 /rpt_family="L1"
repeat_region	1432..1551 /rpt_family="Alu"
repeat_region	1591..1688 /rpt_family="B2"
repeat_region	1862..2137 /rpt_family="B4"
repeat_region	5906..5972 /rpt_family="Alu"
repeat_region	6678..6821 /rpt_family="Alu"
unsure	6940..7110 /note="Unresolved simple sequence repeat."
repeat_region	8021..8354 /rpt_family="MaLR"
repeat_region	8355..10211 /rpt_family="MaLR"
repeat_region	10216..10522 /rpt_family="MaLR"
repeat_region	10726..10844 /rpt_family="B4"
repeat_region	10837..10887 /rpt_family="Alu"
repeat_region	11250..11393

```
/rpt_family="B4"
11606..11705
/rpt_family="B4"
12301..12496
/rpt_family="7SK"
13746..13826
/rpt_family="L2"
15416..15560
/rpt_family="Alu"
15634..15670
/rpt_family="L2"
15798..16052
/rpt_family="MaLR"
16183..16257
/rpt_family="L2"
16289..16570
/rpt_family="B4"
16594..16842
/rpt_family="L2"
18179..18363
/rpt_family="B2"
18371..18459
/rpt_family="L1"
18460..18599
/rpt_family="Alu"
18600..18744
/rpt_family="L1"
18745..18877
/rpt_family="Alu"
18984..19058
/rpt_family="L1"
19079..19131
/rpt_family="MaLR"
19132..19263
/rpt_family="MaLR"
19501..19851
/rpt_family="MaLR"
21000..21039
/rpt_family="L1"
21286..21439
/rpt_family="L1"
21783..22945
/rpt_family="L1"
23027..23224
/rpt_family="B2"
23505..23673
/rpt_family="L1"
23802..23913
/rpt_family="Alu"
24851..24904
/rpt_family="B4"
25986..26296
/rpt_family="MaLR"
27018..27185
/rpt_family="B2"
29142..29242
/rpt_family="Alu"
29508..29606
/rpt_family="Alu"
30880..31954
/note="CpG Island (%GC=72.3, o/e=0.75, #CpGs=116)"
33292..33382
/rpt_family="Alu"
33736..33868
/rpt_family="Alu"
33885..33991
/rpt_family="B2"
34513..34648
/rpt_family="B4"
34799..34953
/rpt_family="B4"
37019..37152
/rpt_family="Alu"

repeat_region 37177..37261
/rpt_family="Alu"
41917..41995
/rpt_family="MIR"
43353..43602
/rpt_family="B4"
44128..44173
/rpt_family="B4"
44174..44526
/rpt_family="MaLR"
44527..44609
/rpt_family="B4"
48077..48211
/rpt_family="MaLR"
48220..48515
/rpt_family="MaLR"
49197..49315
/rpt_family="B4"
49243..49356
/rpt_family="Alu"

Query Match 52.4%; Score 22; DB 6; Length 203589;
Best Local Similarity 83.3%; Pred. No. 1.1e+03;
Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 13 CCGAGCCCTTACAAAGACTATGAGGTAAAG 42
Db 200649 CCAAGCCAGTACAAAGACAATGAGGTAAAG 200620

RESULT 93
AC092090 209153 bp DNA linear HTG 23-MAY-2002
LOCUS Canis familiaris clone RP81-300L5, WORKING DRAFT SEQUENCE, 8
DEFINITION Ordered pieces.
ACCESSION AC092090
VERSION AC092090.2 GI:21104893
KEYWORDS HTG; HTGS PHASE2; HTGS_DRAFT.
SOURCE Canis familiaris (dog)
ORGANISM Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
Canis.
1 (bases 1 to 209153)
Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C.,
Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J.,
Haghighi,P., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E., Laric,P.,
Lee-Lin,S.-Q., Legaepi,R., Maduro,Q.L., Maduro,V.B.,
Margulies,E.H., Masiello,C., Maskeri,B., Mastrian,S.D.,
McCloskey,J.C., McDowell,J., Paguirigan,C., Pearson,S.D.,
Portnoy,M.E., Prasad,A., Schueler,M.G., Stantropop,S., Thomas,J.W.,
Thomas,P.J., Touchman,J.W., Turgeon,C., Vogt,J.L., Walker,M.A.,
Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 209153)
Green,E.D.
Direct Submission
TITLE Submitted (20-JUN-2001) NIH Intramural Sequencing Center, 8717
Grovermont Circle, Gaithersburg, MD 20877, USA
3 (bases 1 to 209153)
Green,E.D.
Direct Submission
TITLE Submitted (23-MAY-2002) NIH Intramural Sequencing Center, 8717
Grovermont Circle, Gaithersburg, MD 20877, USA
On May 23, 2002 this sequence version replaced gi:14495384.
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_zoo@nhgri.nih.gov
----- Project Information
```

Center project name: ayt  
Center clone name: 300L05

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8x average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 207698 bases at least Q40  
Consensus quality: 208254 bases at least Q30  
Consensus quality: 208411 bases at least Q20  
Insert size: 168000; agarose-fp  
Insert size: 208453; sum-of-contigs  
Quality coverage: 13.81x in Q20 bases; agarose-fp  
Quality coverage: 11.13x in Q20 bases; sum-of-contigs  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 8 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\* 1 28791: contig of 28791 bp in length  
\* 28792 28991: gap of unknown length  
\* 28892 94557: contig of 65666 bp in length  
\* 94558 121614: contig of 26957 bp in length  
\* 94658 121715: gap of unknown length  
\* 121615 177060: contig of 55346 bp in length  
\* 121715 177060: gap of unknown length  
\* 177061 180861: contig of 3701 bp in length  
\* 180862 180961: gap of unknown length  
\* 180962 189115: contig of 8254 bp in length  
\* 189115 200836: gap of unknown length  
\* 189216 200836: contig of 11521 bp in length  
\* 200837 209153: contig of 8217 bp in length.  
\* 200937 Location/Qualifiers

#### FEATURES

source  
1..209153  
/organism="Canis familiaris"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9615"  
/clone="RP81-300L5"  
/clone\_lib="RP81"  
1..28791  
/note="assembly\_fragment"  
clone\_end:177  
vector\_side:left  
28792..28891  
/estimated\_length=unknown  
28892..94557  
/note="assembly\_fragment"  
94558..94657  
/estimated\_length=unknown  
94658..121614  
/note="assembly\_fragment"  
121615..121714  
/estimated\_length=unknown  
121715..177060

gap /note="assembly\_fragment"  
177061..177160  
/estimated\_length=unknown  
177161..180861  
/note="assembly\_fragment"  
180862..180961  
/estimated\_length=unknown  
180962..189215  
/note="assembly\_fragment"  
189012..209153  
/note="clone overlaps with GenBank Accession Number  
AC092091 clone RP81-312A9 (center project name ayt)"  
189216..189315  
/estimated\_length=unknown  
189316..200836  
/note="assembly\_fragment"  
200837..200936  
/estimated\_length=unknown  
200937..209153  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:right"  
  
ORIGIN  
  
Query Match 52.4%; Score 22; DB 12; Length 209153;  
Best Local Similarity 73.7%; Pred. No. 1.1e+03;  
Matches 28; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
  
QY 4 CACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAA 41  
Db 133200 CAGGCTTGCTAGCTCTTCCAAAGACTCAGCTGAAA 133237  
  
RESULT 94  
AC111880/c  
AC111880  
LOCUS Rattus norvegicus clone CH230-25K13, WORKING DRAFT SEQUENCE, 3  
DEFINITION Rattus norvegicus clone CH230-25K13, WORKING DRAFT SEQUENCE, 3  
unordered pieces.  
AC111880.5 GI:25006926  
VERSION HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
KEYWORDS Rattus norvegicus (Norway rat)  
SOURCE Rattus norvegicus  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Murioidea; Muridae; Murinae; Rattus.  
1 (bases 1 to 249159)  
REFERENCE  
AUTHORS  
Muzny, D. Marie., Metzker, M. Lee., Abramzon, S., Adams, C., Alder, J.,  
Allen, C., Allen, H., Alsebrook, S., Amin, A., Anguiano, D.,  
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,  
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,  
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,  
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,  
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,  
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,  
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,  
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,  
Egan, A., Escotto, M., Evans, C., Evans, C. A., Falls, T., Fan, G.,  
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerr, W., Guevara, W.,  
Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,  
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,  
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,  
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,  
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,  
Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
Lorensuhera, L., Loulsegged, H., Lozado, R. J., Lu, X., Ma, J.,  
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,





Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,  
 Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,  
 Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,  
 Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,  
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K.,  
 Nwankemeligho, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S.,  
 Pateranak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,  
 Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,  
 Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,  
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S.,  
 Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.,  
 Sanders, W., Savery, G., Scherer, S., Scott, G., Shateman, S., Shen, H.,  
 Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D.,  
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,  
 Steidle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,  
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,  
 Valas, R., Vera, V., Villanana, D., Waldron, L., Walker, B., Wang, J.,  
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,  
 Williams, G., Willison, R., Wleciyk, R., Wooden, H., Worley, K.,  
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von  
 Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,  
 Weinstock, G. and Gibbs, R.A.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 255568)  
 Worley, K.C.  
 Direct Submission  
 Submitted (19-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 255568)  
 Rat Genome Sequencing Consortium.  
 Direct Submission  
 Submitted (19-NOV-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Nov 19, 2002 this sequence version replaced gi:23913093.  
 The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described  
 in the feature table below represents a scaffold in the Atlas  
 assembly (a 'contig-scaffold'). Within each contig-scaffold,  
 individual sequence contigs are ordered and oriented, and separated  
 by sized gaps filled with Ns to the estimated size. The sequence  
 may extend beyond the ends of the clone and there may be sequence  
 contigs within a contig-scaffold that consist entirely of whole  
 genome shotgun sequence reads. Both end sequences and whole genome  
 shotgun sequence only contigs will be indicated in the feature  
 table.

----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
 ----- Project Information  
 Center project name: GZXX  
 Center clone name: CH230-265A10  
 ----- Summary Statistics  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 198395 bases at least Q40  
 Consensus quality: 200298 bases at least Q30  
 Consensus quality: 201736 bases at least Q20  
 Estimated insert size: 205306; sum-of-contigs estimation  
 Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

-----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 4 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1 214174: contig of 214174 bp in length  
 \* 214175 214274: gap of unknown length in length  
 \* 214275 215461: contig of 1187 bp in length  
 \* 215462 215561: gap of unknown length  
 \* 215562 217862: contig of 2301 bp in length  
 \* 217863 217962: gap of unknown length  
 \* 217963 255568: contig of 37606 bp in length.

FEATURES  
 Location/Qualifiers  
 1..255568  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-265A10"  
 1..1265  
 /note="wgs\_end\_extension  
 clone\_end:Sp6"  
 5486..6335  
 /note="clone boundary  
 clone\_end:Sp6"  
 site:  
 end\_sequence:RXAEB05TVC"  
 complement(210537..211391)  
 /note="clone boundary  
 clone\_end:T7"  
 site:  
 end\_sequence:RXAEB05TJC"  
 212221..214174  
 /note="wgs\_end\_extension  
 clone\_end:T7"  
 214175..214274  
 /estimated\_length=unknown  
 215462..215561  
 /estimated\_length=unknown  
 217863..217962  
 /estimated\_length=unknown

ORIGIN  
 Query Match 52.4%; Score 22; DB 12; Length 255568;  
 Best Local Similarity 83.3%; Pred. No. 1.2e+03;  
 Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 12 GCCGAGCCCTTACAAAGACTATGAAGTAA 41  
 |||||  
 Db 234980 GCAGAGCCCTTATATGACTAAGAGCAA 234951  
 |||||

RESULT 96  
 AY844358  
 LOCUS AY844358 428 bp DNA linear VRT 26-AUG-2005  
 DEFINITION Acris crepitans Rag-1 gene, partial cds.  
 ACCESSION AY844358  
 VERSION AY844358.1 GI:61725332  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Acris crepitans (northern cricket frog)  
 Acris crepitans  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Amphibia; Batrachia; Anura; Neobatrachia; Hylidae;  
 Hylinae; Acris.  
 1 (bases 1 to 428)  
 Faivovich, J., Haddad, C.F.B., Garcia, P.C.A., Frost, D.R.,  
 Campbell, J.A. and Wheeler, W.C.  
 Systematic Review of the frog family Hylidae, with special  
 reference to the Hylinae: Phylogenetic analysis and taxonomic  
 revision  
 Bull. Am. Mus. Nat. Hist. 294, 1-240 (2005)  
 2 (bases 1 to 428)  
 Faivovich, J., Haddad, C.F.B., Garcia, P.C.A., Frost, D.R.,  
 Campbell, J.A. and Wheeler, W.C.  
 Direct Submission  
 TITLE

JOURNAL Submitted (01-DEC-2004) Herpetology, American Museum of Natural History, Central Park West at 79th Street, New York, NY 10024, USA

FEATURES

1 source

Location/Qualifiers

1. .428

/organism="Acris crepitans"

/mol\_type="genomic DNA"

/specimen\_voucher="LSUMZ H-2164"

/db\_xref="taxon:317300"

/country="USA: Alabama, De Kalb Co., Powerline access Rd., 1/10 mi W of Lookout Mt. Boys Camp Rd."

<1. .>428

/product="Rag-1"

<1. .>428

/codon\_start=3

/product="Rag-1"

/protein\_id="AA54068.1"

/db\_xref="GI:61725335"

/translation="PIEERKKWOATLDKHLRKKMNLKPIRMNNGNFARKLMKSTVEA VCDLIPSEERQAALRELDLYLKMFPVWRSSCPAKECPPELLCQYSYHSORFAELLSTK FKRYEGKITNYFHKTLAHVPEIIERDGSIGAWASEGNES"

ORIGIN

Query Match 51.9%; Score 21.8; DB 11; Length 428;

Best Local Similarity 70.7%; Pred. No. 1.6e+02;

Matches 29; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1 GCCACGCTTGCCGAGCCCTTACAAGACTATGAAGTAAA 41

|||||

Db 29 GGCCACCCCTTGACAAACACCTTAGAAGAAGATGAACCTTAA 69

|||||

RESULT 97

AY844359

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Acris gryllus (Florida cricket frog)

REFERENCE

1 (bases 1 to 428)

FAIVOVICH,J., HADDAD,C.F.B., GARCIA,P.C.A., FROST,D.R., CAMPBELL,J.A. and WHEELER,W.C.

Systematic Review of the frog family Hylidae, with special reference to the Hylinae: Phylogenetic analysis and taxonomic revision

Bull. Am. Mus. Nat. Hist. 294, 1-240 (2005)

2 (bases 1 to 428)

FAIVOVICH,J., HADDAD,C.F.B., GARCIA,P.C.A., FROST,D.R., CAMPBELL,J.A. and WHEELER,W.C.

Direct Submission

Submitted (01-DEC-2004) Herpetology, American Museum of Natural History, Central Park West at 79th Street, New York, NY 10024, USA

FEATURES

source

1. .428

/organism="Acris gryllus"

/mol\_type="genomic DNA"

/db\_xref="taxon:168422"

/country="USA: Florida, Okaloosa Co., Eglin AFB, JR Walton Pond. 30 41.59'N 86 28.36'W"

<1. .>428

/product="Rag-1"

<1. .>428

/codon\_start=3

/product="Rag-1"

/protein\_id="AA54069.1"

/db\_xref="GI:61725335"

/translation="PIEERKKWOATLDKHLRKKMNLKPIRMNNGNFARKLMKSTVEA VCDLIPSEERQAALRELDLYLKMFPVWRSSCPAKECPPELLCQYSYHSORFAELLSTK FKRYEGKITNYFHKTLAHVPEIIERDGSIGAWASEGNES"

ORIGIN

Query Match 51.9%; Score 21.8; DB 11; Length 428;

Best Local Similarity 70.7%; Pred. No. 1.6e+02;

Matches 29; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1 GCCACGCTTGCCGAGCCCTTACAAGACTATGAAGTAAA 41

|||||

Db 29 GGCCACCCCTTGACAAACACCTTAGAAGAAGATGAACCTTAA 69

|||||

RESULT 99

AY844391

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Hyla arborea (European tree frog)

REFERENCE

1 (bases 1 to 428)

FAIVOVICH,J., HADDAD,C.F.B., GARCIA,P.C.A., FROST,D.R., CAMPBELL,J.A. and WHEELER,W.C.

Systematic Review of the frog family Hylidae, with special reference to the Hylinae: Phylogenetic analysis and taxonomic revision

Bull. Am. Mus. Nat. Hist. 294, 1-240 (2005)

2 (bases 1 to 428)

FAIVOVICH,J., HADDAD,C.F.B., GARCIA,P.C.A., FROST,D.R., CAMPBELL,J.A. and WHEELER,W.C.

Direct Submission

Submitted (01-DEC-2004) Herpetology, American Museum of Natural History, Central Park West at 79th Street, New York, NY 10024, USA

FEATURES

source

1. .428

/organism="Hyla arborea"

/mol\_type="genomic DNA"

/db\_xref="taxon:121869"

/country="Germany: Petrade"

<1. .>428

/product="Rag-1"

<1. .>428

/codon\_start=3

/product="Rag-1"

/protein\_id="AA54099.1"

/db\_xref="GI:61725395"

/translation="FVEERKKWOATLDKHLRKKMNLKPIRMNNGNFARKLMKSTVEA VCDLIPSEERQAALRELDLYLKMFPVWRSSCPAKECPPELLCQYSYHSORFAELLSTK FKRYEGKITNYFHKTLAHVPEIIERDGSIGAWASEGNES"

ORIGIN

Query Match 51.9%; Score 21.8; DB 11; Length 428;

Best Local Similarity 70.7%; Pred. No. 1.6e+02;

Matches 29; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1 GCCACGCTTGCCGAGCCCTTACAAGACTATGAAGTAAA 41

|||||

Db 29 GGCAACGCTTGACAAACACCTTAGAAGAAGATGAACCTTAA 69

|||||

RESULT 99

AY844391

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Hyla arenicolor Rag-1 gene, partial cds.

REFERENCE

1 (bases 1 to 428)

FAIVOVICH,J., HADDAD,C.F.B., GARCIA,P.C.A., FROST,D.R., CAMPBELL,J.A. and WHEELER,W.C.

Direct Submission

Submitted (01-DEC-2004) Herpetology, American Museum of Natural History, Central Park West at 79th Street, New York, NY 10024, USA

FEATURES

source

1. .428

/organism="Hyla arenicolor"

/mol\_type="genomic DNA"

/db\_xref="taxon:121869"

/country="Germany: Petrade"

<1. .>428

/product="Rag-1"

<1. .>428

/codon\_start=3

/product="Rag-1"

/protein\_id="AA54099.1"

/db\_xref="GI:61725395"

/translation="FVEERKKWOATLDKHLRKKMNLKPIRMNNGNFARKLMKSTVEA VCDLIPSEERQAALRELDLYLKMFPVWRSSCPAKECPPELLCQYSYHSORFAELLSTK FKRYEGKITNYFHKTLAHVPEIIERDGSIGAWASEGNES"

ORIGIN

Query Match 51.9%; Score 21.8; DB 11; Length 428;

Best Local Similarity 70.7%; Pred. No. 1.6e+02;

Matches 29; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1 GCCACGCTTGCCGAGCCCTTACAAGACTATGAAGTAAA 41

|||||

Db 29 GGCAACGCTTGACAAACACCTTAGAAGAAGATGAACCTTAA 69

|||||

RESULT 99

AY844391

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Hyla arenicolor Rag-1 gene, partial cds.

REFERENCE

1 (bases 1 to 428)

FAIVOVICH,J., HADDAD,C.F.B., GARCIA,P.C.A., FROST,D.R., CAMPBELL,J.A. and WHEELER,W.C.

Direct Submission

Submitted (01-DEC-2004) Herpetology, American Museum of Natural History, Central Park West at 79th Street, New York, NY 10024, USA

FEATURES

source

1. .428

/organism="Hyla arenicolor"

/mol\_type="genomic DNA"

/db\_xref="taxon:121869"

/country="Germany: Petrade"

<1. .>428

/product="Rag-1"

<1. .>428

/codon\_start=3

/product="Rag-1"

/protein\_id="AA54099.1"

/db\_xref="GI:61725395"

/translation="FVEERKKWOATLDKHLRKKMNLKPIRMNNGNFARKLMKSTVEA VCDLIPSEERQAALRELDLYLKMFPVWRSSCPAKECPPELLCQYSYHSORFAELLSTK FKRYEGKITNYFHKTLAHVPEIIERDGSIGAWASEGNES"

ORIGIN

Query Match 51.9%; Score 21.8; DB 11; Length 428;

Best Local Similarity 70.7%; Pred. No. 1.6e+02;

Matches 29; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1 GCCACGCTTGCCGAGCCCTTACAAGACTATGAAGTAAA 41

|||||

Db 29 GGCAACGCTTGACAAACACCTTAGAAGAAGATGAACCTTAA 69

|||||

RESULT 99

AY844391

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Hyla arenicolor Rag-1 gene, partial cds.

REFERENCE

1 (bases 1 to 428)

FAIVOVICH,J., HADDAD,C.F.B., GARCIA,P.C.A., FROST,D.R., CAMPBELL,J.A. and WHEELER,W.C.

Direct Submission

Submitted (01-DEC-2004) Herpetology, American Museum of Natural History, Central Park West at 79th Street, New York, NY 10024, USA

FEATURES

source

1. .428

/organism="Hyla arenicolor"

/mol\_type="genomic DNA"

/db\_xref="taxon:121869"

/country="Germany: Petrade"

<1. .>428

/product="Rag-1"

<1. .>428

/codon\_start=3

/product="Rag-1"

/protein\_id="AA54099.1"

/db\_xref="GI:61725395"

/translation="FVEERKKWOATLDKHLRKKMNLKPIRMNNGNFARKLMKSTVEA VCDLIPSEERQAALRELDLYLKMFPVWRSSCPAKECPPELLCQYSYHSORFAELLSTK FKRYEGKITNYFHKTLAHVPEIIERDGSIGAWASEGNES"

ORIGIN

Query Match 51.9%; Score 21.8; DB 11; Length 428;

Best Local Similarity 70.7%; Pred. No. 1.6e+02;

Matches 29; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1 GCCACGCTTGCCGAGCCCTTACAAGACTATGAAGTAAA 41

|||||

Db 29 GGCAACGCTTGACAAACACCTTAGAAGAAGATGAACCTTAA 69

|||||



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2006, 15:16:40 ; Search time 312.2 Seconds

(without alignments)  
937.971 Million cell updates/sec

Title: US-10-600-816-30

Perfect score: 42

Sequence: 1 gccacgcttgccgagccc.....acaaagactatgaagtaag 42

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 5244920 seqs, 3486124231 residues

Total number of hits satisfying chosen parameters: 10489840

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum First 100%

Listing first 1000 summaries

Database : N Geneseq 8:\*

1: geneseqn1980s:\*

2: geneseqn1990s:\*

3: geneseqn2000s:\*

4: geneseqn2001as:\*

5: geneseqn2001bs:\*

6: geneseqn2002as:\*

7: geneseqn2002bs:\*

8: geneseqn2003as:\*

9: geneseqn2003bs:\*

10: geneseqn2003cs:\*

11: geneseqn2003ds:\*

12: geneseqn2004as:\*

13: geneseqn2004bs:\*

14: geneseqn2005s:\*

15: geneseqn2006s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	ID	Description
1	42	100.0	497 11	ADT96418 Colon can
2	42	100.0	497 11	ADX42900 Human CDN
3	42	100.0	552 11	ADT96292 Colon can
4	42	100.0	552 11	ADX42774 Human CDN
5	42	100.0	603 3	AAZ90046 Human col
6	42	100.0	620 6	ABQ58527 Human col
7	42	100.0	634 6	ABQ59698 Human col
8	42	100.0	642 10	ABT22328 Breast ca
9	42	100.0	1074 12	ADG30035 Human GPC
10	42	100.0	1212 3	AAZ61776 cDNA enco
11	42	100.0	1212 4	AAC99709 Skin cell
12	42	100.0	1212 6	ABL34861 Human CDN
13	42	100.0	1228 8	ACA04775 cDNA enco
14	42	100.0	1460 14	AEA00112 Human TAT
15	42	100.0	1460 14	AEA00632 Human TAT
16	42	100.0	1619 4	AAP58615 Human REC
17	42	100.0	1718 3	AAZ90056 Hydrophob
18	42	100.0	2302 6	ABT10173 Human bre

Acc58386 Human GPC  
Add93240 RAIG1 cod  
Adi24773 Intestina  
Abt31923 Human bre  
Ahl4688 Human cDN  
Abz42832 Human G p  
Acc72695 Human can  
Adn39301 Cancer/an  
Adi28525 Human GPC  
Adi28459 Human GPC  
Adi48221 Human ret  
Acr39388 Tumour-as  
Acr43992 Human bre  
Adu06126 Novel bro  
Aea00088 Human TAT  
Aea00608 Human TAT  
Aed47480 Retinoloic  
Abq54954 Human ova  
Aah72766 Human cer  
Acr89274 Breast ca  
Adx97494 Pancreat  
Adf70573 Orphan re  
Aah08354 Human CDN  
Adi28530 Human GPC  
Adi28527 Human GPC  
Adv92548 Japanese  
Aaa29595 Japanese  
Adv92550 Japanese  
Adv92562 Japanese  
Adm97133 Japanese-  
Aaq22767 JEV Nakay  
Aan80297 Partial n  
Adw74114 PCJEME DN  
Abk62169 Rat seque  
Adb55155 Toxicity-  
Adb49676 Primary r  
Adc85956 Human GPC  
Abx63126 Human CDN  
Adj56498 Human CDN  
Abk42428 Genomic s  
Aal07451 Human rep  
Adb60584 Connectiv  
Abk42426 Genomic s  
Aal07450 Human rep  
Adb60582 Connectiv  
Abi17076 Drosophil  
Aed85840 Human CD3  
Aed85870 Human CD3  
Aac16864 Human sec  
Aad42934 Human pho  
Aac28686 Human sec  
Abi03149 Drosophil  
Aaw52673 Japanese  
Aas92414 DNA encod  
Adc60894 Bacterial  
Adi70879 Human ova  
Adi71933 Human ova  
Adi73027 Human ova  
Adi38162 Human ova  
Adi36609 Human ova  
Adi71451 Human ova  
Adi44546 Human ova  
Adi44897 Human ova  
Ado30325 Mouse GPC  
Adi62421 Human ova  
Adi62598 Human ova  
Adi10225 Human the  
Acr38091 Tumour-as  
Abn59642 Novel hum  
Adi62578 Human ova  
Adi11553 Human the  
Adi22610 Human liv

92	21	50.0	6346	12	ADQ22756	Adq22756 Human sof	165	20	47.6	667	13	ADR62740	Adr62740 Cotton cd
93	21	50.0	1714	10	ADC37602	Adc37602 Human nuc	c 166	20	47.6	1123	6	ABL89772	AbL89772 Human pol
c 94	21	50.0	96589	9	ADA02708	Ada02708 Human ZFH	c 167	20	47.6	7690	6	ABL33123	AbL33123 Human imm
c 95	21	50.0	96589	10	ADB72446	Adb72446 Human ZFH	c 168	20	47.6	10591	12	ADO59810	Ado59810 Novel hum
c 96	21	50.0	96589	10	ADE95956	Ade95956 Human ZFH	c 169	20	47.6	38703	14	AED83346	Aed83346 Human cos
c 97	20.8	49.5	1728	10	ADC93459	Adc93459 E. faeciu	c 170	19.8	47.1	546	3	AAF08124	Aaf08124 Fusarium
c 98	20.6	49.0	377	2	ADT23493	Adt23493 Human gen	c 171	19.8	47.1	546	13	ADU52165	Adu52165 Fusarium
c 99	20.6	49.0	457	9	ACH22183	Ach22183 Human adu	c 172	19.8	47.1	546	14	ADZ90168	Adz90168 Fusarium
c 100	20.6	49.0	682	5	ADL45933	Adl45933 Human ova	c 173	19.8	47.1	726	6	ABN76605	Abn76605 Human syn
c 101	20.6	49.0	934	13	ADX60641	Adx60641 Plant ful	c 174	19.8	47.1	1125	13	ADT45575	Adt45575 Bacterial
c 102	20.6	49.0	963	4	AAF33117	Aaf33117 Human sec	c 175	19.8	47.1	1128	13	ADS46516	Ads46516 Bacterial
c 103	20.6	49.0	1764	3	AACT77933	Aac77933 Human can	c 176	19.8	47.1	1405	6	AAK99775	Aak99775 1405st DN
c 104	20.6	49.0	2085	5	ABV28029	Abv28029 Human pro	c 177	19.8	47.1	1545	10	AAAF90206	Aaf90206 Nucleotid
c 105	20.6	49.0	2085	5	ABV22192	Abv22192 Human pro	c 178	19.8	47.1	1545	10	ADB69816	Adb69816 C. neofor
c 106	20.6	49.0	2085	5	ABV27935	Abv27935 Human pro	c 179	19.8	47.1	1720	8	ABX63106	Abx63106 Human CDN
c 107	20.6	49.0	2085	5	ABV22095	Abv22095 Human pro	c 180	19.8	47.1	6683	10	ADB53086	Adb53086 Primary r
c 108	20.6	49.0	2387	13	ACN40823	Acn40823 Tumour-as	c 181	19.8	47.1	6683	13	ADVA1291	Adv41291 Rat cardi
c 109	20.6	49.0	2510	3	AAZ90578	Aaz90578 Human dea	c 182	19.8	47.1	6874	4	AAAS45441	Aas45441 Chemicall
c 110	20.6	49.0	2759	13	ACN40821	Acn40821 Tumour-as	c 183	19.8	47.1	6874	4	ABL70382	AbL70382 Chemicall
c 111	20.6	49.0	2759	14	ABEB2826	Abb2826 Human col	c 184	19.8	47.1	6874	6	ABK28290	Abk28290 DNA trans
c 112	20.6	49.0	2772	4	AAH15102	Aah15102 Human CDN	c 185	19.8	47.1	6874	6	AAAS61334	Aas61334 Human gen
c 113	20.6	49.0	2895	13	ADU01853	Adu01853 Novel hum	c 186	19.8	47.1	6874	6	ABN80231	Abn80231 Human che
c 114	20.6	49.0	2977	13	ACN40822	Acn40822 Tumour-as	c 187	19.8	47.1	10699	4	ACN14610	Acn14610 Wild-type
c 115	20.6	49.0	2977	14	ABEB2793	Abb2793 Human col	c 188	19.8	47.1	10699	4	ABD14611	Abd14611 Attenuate
c 116	20.6	49.0	4279	14	ADX07488	Adx07488 Cyclin-de	c 189	19.8	47.1	11660	6	ABN85368	Abn85368 Spumaret
c 117	20.6	49.0	4506	15	ABE87869	Abe87869 Human CDN	c 190	19.8	47.1	143899	6	AAU38336	Aau38336 Genomic s
c 118	20.6	49.0	4522	5	AAAS64864	Aas64864 DNA encod	c 191	19.6	46.7	49	14	ADV68377	Adv68377 Heyvea bra
c 119	20.6	49.0	6579	8	ABT18965	Abt18965 Aspergill	c 192	19.6	46.7	272	12	ADQ06087	Adq06087 Soybean t
c 120	20.6	49.0	7290	4	AAK53352	Aak53352 Human pol	c 193	19.6	46.7	337	8	AAVF78634	Aavf78634 Staphyloc
c 121	20.6	49.0	7838	4	AAK52368	Aak52368 Human pol	c 194	19.6	46.7	405	8	ACF73285	Acf73285 Staphyloc
c 122	20.6	49.0	30826	4	AAK67051	Aak67051 Human imm	c 195	19.6	46.7	539	10	ADD34794	Add34794 Mouse nit
c 123	20.4	48.6	264	5	AAF68875	Aaf68875 Human lun	c 196	19.6	46.7	692	13	ADX36505	Adx36505 Plant ful
c 124	20.4	48.6	264	6	ABK38786	Abk38786 cDNA clon	c 197	19.6	46.7	750	8	ACF64343	Acf64343 Human IFN
c 125	20.4	48.6	264	8	ACA11115	Aca11115 Human lun	c 198	19.6	46.7	879	6	ABS51668	Abs51668 LEC-1-rel
c 126	20.4	48.6	264	8	ACA02301	Aca02301 Lung canc	c 199	19.6	46.7	879	10	ADC23662	Acd23662 cDNA enco
c 127	20.4	48.6	264	10	ADH46342	Adh46342 Human lun	c 200	19.6	46.7	880	8	ACC00774	Acc00774 Glycine m
c 128	20.4	48.6	264	13	ADU20261	Adj20261 Human lun	c 201	19.6	46.7	946	13	ADX27655	Adx27655 Plant ful
c 129	20.4	48.6	363	14	ACL67473	Adl67473 M. xanthu	c 202	19.6	46.7	965	8	ACD20405	Acd20405 DNA encod
c 130	20.4	48.6	629	5	ABV55696	Abv55696 Human pro	c 203	19.6	46.7	988	2	AAQ70155	Aaq70155 Sequence
c 131	20.4	48.6	707	6	ABK78003	Abk78003 Bacillus	c 204	19.6	46.7	993	12	ADI42284	Adi42284 Plant tra
c 132	20.4	48.6	708	6	ABK77843	Abk77843 Bacillus	c 205	19.6	46.7	993	12	ADO02743	Ado02743 Soybean o
c 133	20.4	48.6	728	4	AAH06438	Aah06438 Human CDN	c 206	19.6	46.7	993	12	ADO62332	Ado62332 Transcript
c 134	20.4	48.6	1659	13	ADT87784	Adt87784 Plant CDN	c 207	19.6	46.7	1020	3	AAA49789	Aaa49789 Arabidops
c 135	20.4	48.6	2318	4	AAK53300	Aak53300 Human pol	c 208	19.6	46.7	1222	2	AAQ70158	Aaq70158 Sequence
c 136	20.4	48.6	2523	10	ADA53709	Ada53709 Human cod	c 209	19.6	46.7	1344	8	ACA37293	Aca37293 Prokaryot
c 137	20.4	48.6	2623	4	AAK52316	Aak52316 Human pol	c 210	19.6	46.7	1428	10	ABT42343	Abt42343 Toxicity
c 138	20.4	48.6	2648	10	ADB62195	Adb62195 Human CDN	c 211	19.6	46.7	1428	13	ADVA1584	Adv41584 Rat cardi
c 139	20.4	48.6	2737	4	ABU24858	Abu24858 Drosophil	c 212	19.6	46.7	1500	1	AAW71394	Aaw71394 Flaviviru
c 140	20.4	48.6	3057	15	AAE88935	Aae88935 Human che	c 213	19.6	46.7	1531	2	AAQ70156	Aaq70156 Sequence
c 141	20.4	48.6	3240	4	AAH14453	Aah14453 Human CDN	c 214	19.6	46.7	2040	4	AAH26533	Aah26533 Human pro
c 142	20.4	48.6	21295	14	ACL64731	Adl64731 M. xanthu	c 215	19.6	46.7	2054	13	ADW23987	Adw23987 Japanese
c 143	20.4	48.6	122614	11	ACN44998	Acn44998 Human gen	c 216	19.6	46.7	2070	4	AAH26532	Aah26532 Human pro
c 144	20.4	48.6	167932	10	ADL13501	Adl13501 Osteoarth	c 217	19.6	46.7	2771	15	ABA18382	Abal8382 Human ner
c 145	20.4	48.6	265118	5	AAH41227	Aah41227 Pyrococcu	c 218	19.6	46.7	3110	13	ADW23445	Adw23445 Japanese
c 146	20.2	48.1	384	6	ABES4351	Abes4351 DNA encod	c 219	19.6	46.7	3577	6	ABI99459	Abi99459 Mouse isc
c 147	20.2	48.1	614	3	AAAC45316	Aac45316 Arabidops	c 220	19.6	46.7	5293	10	ABV77549	Abv77549 Plasmid p
c 148	20.2	48.1	1410	8	ACAA1716	Aca41716 Prokaryot	c 221	19.6	46.7	5334	10	ABV77537	Abv77537 Japanese
c 149	20.2	48.1	1419	3	AAAC44653	Aac44653 Arabidops	c 222	19.6	46.7	7500	10	ABV77533	Abv77533 Japanese
c 150	20.2	48.1	1545	6	AAAS97238	Aas97238 Neisseria	c 223	19.6	46.7	9310	2	AAV74627	Aav74627 Staphyloc
c 151	20.2	48.1	1548	3	AA254374	Aa254374 Neisseria	c 224	19.6	46.7	10892	4	AAF83821	Aaf83821 Chimeric
c 152	20.2	48.1	5587	10	ABES5229	Abe55229 Rat gene	c 225	19.6	46.7	10892	4	AAED66292	Aed66292 Yellow fe
c 153	20.2	48.1	11851	12	ADL98344	Adl98344 lin(in)3628	c 226	19.6	46.7	10976	3	ABL50890	AbL50890 Japanese
c 154	20.2	48.1	108359	9	ADA13316	Adal13316 Human fri	c 227	19.6	46.7	10985	4	AAK86420	Aak86420 Human imm
c 155	20	47.6	489	13	ACN57177	Acn57177 Cotton gy	c 228	19.6	46.7	10985	4	AAK86422	Aak86422 Human imm
c 156	20	47.6	527	4	AAI16854	Aai16854 Human bre	c 229	19.6	46.7	10985	4	AAK86421	Aak86421 Human imm
c 157	20	47.6	531	6	AAK98730	Aak98730 DNA seque	c 230	19.6	46.7	41038	14	ABE96523	Abe96523 Human IFN
c 158	20	47.6	542	4	AAAD03843	Aad03843 Complemen	c 231	19.6	46.7	51558	13	ACN37207	Acn37207 Human per
c 159	20	47.6	542	4	AAAD03842	Aad03842 FLAG-huma	c 232	19.6	46.7	52677	11	ACN43934	Acn43934 Human gen
c 160	20	47.6	546	6	AAK98733	Aak98733 DNA of a	c 233	19.6	46.7	87320	14	ADZ12863	Adz12863 Human can
c 161	20	47.6	576	2	AAQ83893	Aaq83893 Hepatitis	c 234	19.6	46.7	110000	14	ABE39172	AbE39172 Continuation (2 of
c 162	20	47.6	576	2	AAAT16607	Aat16607 Hepatitis	c 235	19.6	46.7	110000	14	ABE42736	AbE42736 Continuation (5 of
c 163	20	47.6	608	13	ACN61430	Acn61430 Cotton gy	c 236	19.6	46.7	349980	6	ABQ81849	Abq81849 Bifidobac
c 164	20	47.6	649	5	ADL42993	Adl42993 Human ova	c 237	19.4	46.2	480	13	ACF87190	Acf87190 Human SIR

C 238	19.4	46.2	546	12	ADJ10916	Adi10916 Recombina	C 311	19.2	45.7	2325	8	ACC46631	Acc46631 Human dit
C 239	19.4	46.2	570	5	ADL33095	Adl33095 Human ova	C 312	19.2	45.7	2428	4	AAH98381	Aah98381 Human EST
C 240	19.4	46.2	810	6	ABL06697	Abi106697 Human dit	C 313	19.2	45.7	2428	4	AAH98396	Aah98396 Human EST
C 241	19.4	46.2	844	4	ABL06697	Abi106697 Human dit	C 314	19.2	45.7	2451	2	AAV15809	Aav15809 Microocc
C 242	19.4	46.2	883	10	ADBS8936	Adb58936 Toxicity-	C 315	19.2	45.7	3144	11	ADM03175	Adm03175 Human cDN
C 243	19.4	46.2	883	10	ABT41733	Abt41733 Toxicity	C 316	19.2	45.7	3144	14	ASE86105	Aec86105 Human cDN
C 244	19.4	46.2	883	12	ADP72564	Adp72564 Renal tox	C 317	19.2	45.7	3257	3	AAZ52447	Aaz52447 HTRM clon
C 245	19.4	46.2	1682	8	ACA22852	Acas22852 Prokaryot	C 318	19.2	45.7	3500	13	ADR07983	Adr07983 Full leng
C 246	19.4	46.2	1725	3	AAA96701	Aaa96701 Reporayot	C 319	19.2	45.7	3507	14	ADX07941	Adx07941 Cyclin-de
C 247	19.4	46.2	2173	4	ABL06696	Abi106696 Drosophill	C 320	19.2	45.7	3507	14	ADZ49474	Adz49474 Insulin s
C 248	19.4	46.2	2425	5	AAS79360	Aas79360 DNA encod	C 321	19.2	45.7	3675	4	ABL14979	Abi14979 Drosophill
C 249	19.4	46.2	2499	10	ADF28508	Adf28508 NgrHy con	C 322	19.2	45.7	4005	8	ABT18026	Abt18026 Aspergill
C 250	19.4	46.2	2565	10	ADE06964	Ade06964 Novel cod	C 323	19.2	45.7	4235	8	ABT19840	Abt19840 Aspergill
C 251	19.4	46.2	3142	4	ABL14842	Abi14842 Drosophill	C 324	19.2	45.7	4300	4	AAH42107	Aah42107 Nucleotid
C 252	19.4	46.2	3750	15	AEE26295	Aee26295 BRCT doma	C 325	19.2	45.7	4909	14	AEA27507	Aea27507 Human zin
C 253	19.4	46.2	3895	11	ACN44011	Acn44011 Human mRN	C 326	19.2	45.7	4909	14	AEA30110	Aea30110 Glioblast
C 254	19.4	46.2	6878	4	ABL17536	Abi17536 Drosophill	C 327	19.2	45.7	4960	6	ABS73413	Abs73413 Human G11
C 255	19.4	46.2	7430	4	AAK77418	Aak77418 Human imm	C 328	19.2	45.7	5191	10	ADH28900	Adh28900 Human chr
C 256	19.4	46.2	7894	8	ABZ76334	Abz76334 L. cuprin	C 329	19.2	45.7	5191	12	ADG39645	Adg39645 Human cDN
C 257	19.4	46.2	10002	8	ABZ76333	Abz76333 L. cuprin	C 330	19.2	45.7	5191	12	ADG39694	Adg39694 Human cDN
C 258	19.4	46.2	10341	8	ABZ76340	Abz76340 L. cuprin	C 331	19.2	45.7	5264	6	AAZ37432	Aaz37432 PCDD2/acc
C 259	19.4	46.2	20210	4	AAK77417	Aak77417 Human imm	C 332	19.2	45.7	5490	6	ABQ99289	Abq99289 Human cod
C 260	19.4	46.2	20210	4	AAS31519	Aas31519 Human DNA	C 333	19.2	45.7	5906	15	AEF74614	Aef74614 Human pol
C 261	19.4	46.2	20210	6	ABQ66843	Abq66843 Human pol	C 334	19.2	45.7	5981	8	ABX63291	Abx63291 Human cDN
C 262	19.4	46.2	20210	10	ADC11130	Adc11130 Human DNA	C 335	19.2	45.7	7619	4	ABL14978	Abi14978 Drosophill
C 263	19.4	46.2	38019	14	AEE04707	Aee04707 Cancer-as	C 336	19.2	45.7	9027	8	AAH49171	Aah49171 CFT1030 v
C 264	19.4	46.2	83836	11	ACN45080	Acn45080 Mouse gen	C 337	19.2	45.7	9259	8	AAH49174	Aah49174 CFT1030 v
C 265	19.4	46.2	108526	14	ADZ13294	Adz13294 Murine ca	C 338	19.2	45.7	10553	8	AAH49160	Aah49160 CFT710 ve
C 266	19.4	46.2	110000	14	AEA61120_4	Continuation (5 of	C 339	19.2	45.7	38358	4	AAK73535	Aak73535 Human imm
C 267	19.4	46.2	110000	14	AEA61120_2	Continuation (3 of	C 340	19.2	45.7	106323	8	ABX14591	Abx14591 Human chl
C 268	19.4	46.2	119854	14	ADZ13282	Adz13282 Murine ca	C 341	19.2	45.7	106323	10	AAD58821	Aad58821 Human tra
C 269	19.4	46.2	157090	12	ADO47194	Ado47194 DNA seque	C 342	19.2	45.7	107159	12	ADL13431	Adl13431 Human thy
C 270	19.4	46.2	196686	11	ACN44170	Acn44170 Human gen	C 343	19.2	45.7	110000	10	ACF67367_22	Continuation (23 o
C 271	19.4	46.2	198522	11	ACN44010	Acn44010 Human gen	C 344	19.2	45.7	110000	10	ACF67367_23	Continuation (24 o
C 272	19.4	46.2	325093	14	AEE25355	Aee25355 Human ARR	C 345	19.2	45.7	110000	10	ACF65386_5	Continuation (6 of
C 273	19.2	45.7	252	5	ABV07221	Abv07221 Human pro	C 346	19.2	45.7	139308	8	ADB12769	Adb12769 Human PRK
C 274	19.2	45.7	395	4	AAI84093	Aai84093 Human pol	C 347	19.2	45.7	155083	12	ADQ97891	Adq97891 Mouse can
C 275	19.2	45.7	398	5	ABV27161	Abv27161 Human pro	C 348	19.2	45.7	202251	11	ACN44504	Acn44504 Mouse gen
C 276	19.2	45.7	450	11	ABD09803	Abd09803 Pseudomon	C 349	19.2	45.7	202251	12	ADL128529	Adl128529 Human GPC
C 277	19.2	45.7	459	10	ACF69385	Acf69385 Pseudorhab	C 350	19.2	45.7	289	6	ABL17178	Abi17178 Corn tass
C 278	19.2	45.7	465	4	AAI15537	Aai15537 Probe #54	C 351	19.2	45.7	297	9	ACH40353	Ach40353 Human ant
C 279	19.2	45.7	465	4	ABA57440	Abas7440 Human foe	C 352	19.2	45.7	456	9	ACH40353	Ach40353 Human foe
C 280	19.2	45.7	465	4	AAI36988	Aai36988 Probe #56	C 353	19.2	45.7	457	13	ACF84030	Acf84030 Human STR
C 281	19.2	45.7	465	4	ABA26945	Abas26945 Probe #54	C 354	19.2	45.7	471	10	ADE59600	Ades59600 Rat gene
C 282	19.2	45.7	465	4	AAK31079	Aak31079 Human bon	C 355	19.2	45.7	482	9	ACH41862	Ach41862 Human foe
C 283	19.2	45.7	465	4	AAK30758	Aak30758 Human liv	C 356	19.2	45.7	489	5	AAH81814	Aah81814 Human dif
C 284	19.2	45.7	465	4	ABK05478	Abk05478 Human bra	C 357	19.2	45.7	489	5	AAH81814	Aah81814 Human dif
C 285	19.2	45.7	465	6	ABS05830	Abs05830 Human Gen	C 358	19.2	45.7	514	8	ABZ53921	Abz53921 Aspergill
C 286	19.2	45.7	507	13	ADP63029	Adp63029 Cotton cD	C 359	19.2	45.7	529	13	ACF91105	Acf91105 Human STR
C 287	19.2	45.7	578	13	ACN49195	Acn49195 Cotton pr	C 360	19.2	45.7	536	13	ADQ54362	Adq54362 Novel can
C 288	19.2	45.7	579	2	AAQ78092	Aaq78092 Hepatitis	C 361	19.2	45.7	566	4	AAH13627	Aah13627 Human cDN
C 289	19.2	45.7	579	2	AAQ78113	Aaq78113 Hepatitis	C 362	19.2	45.7	576	2	AAQ83888	Aaq83888 Hepatitis
C 290	19.2	45.7	579	2	AAQ78093	Aaq78093 Hepatitis	C 363	19.2	45.7	576	2	AAQ16602	Aat16602 Hepatitis
C 291	19.2	45.7	580	2	AAQ78045	Aaq78045 Hepatitis	C 364	19.2	45.7	1188	2	AAQ22974	Aaq22974 Sequence
C 292	19.2	45.7	580	2	AAQ78046	Aaq78046 Hepatitis	C 365	19.2	45.7	1224	13	ADR63445	Adr63445 Cotton cD
C 293	19.2	45.7	630	2	AAI12966	Aai12966 HCV El co	C 366	19.2	45.7	1260	2	AAQ22972	Aaq22972 Sequence
C 294	19.2	45.7	630	6	AAI489930	Aai489930 Hepatitis	C 367	19.2	45.7	1263	14	ADV42778	Adv42778 Human psy
C 295	19.2	45.7	630	10	ADD55539	Add55539 Hepatitis	C 368	19.2	45.7	1327	3	AAA34861	Aaa34861 Human low
C 296	19.2	45.7	630	12	ADP71121	Adp71121 HCV DNA e	C 369	19.2	45.7	1327	3	AAQ20983	Aaf20983 Human low
C 297	19.2	45.7	959	2	AAQ78049	Aaq78049 Hepatitis	C 370	19.2	45.7	1327	10	ABZ96677	Abz96677 Human nuc
C 298	19.2	45.7	959	2	AAQ78047	Aaq78047 Hepatitis	C 371	19.2	45.7	1327	11	ABD20526	Abd20526 Human pul
C 299	19.2	45.7	959	2	AAQ78048	Aaq78048 Hepatitis	C 372	19.2	45.7	1351	2	AAQ25789	Aaq25789 Encodes s
C 300	19.2	45.7	1311	13	ADP63027	Adp63027 Cotton cD	C 373	19.2	45.7	1351	2	AAQ38440	Aaq38440 SHL-SR-a
C 301	19.2	45.7	1491	3	AACT76674	Aact76674 Human ORF	C 374	19.2	45.7	1372	11	AEA90313	Aea90313 Human NOV
C 302	19.2	45.7	1641	8	ABT19214	Abt19214 Aspergill	C 375	19.2	45.7	1405	3	AAA34864	Aaa34864 Human low
C 303	19.2	45.7	1683	11	ABD09733	Abd09733 Pseudomon	C 376	19.2	45.7	1405	3	AAQ20986	Aaf20986 Human low
C 304	19.2	45.7	1725	13	ADS14550	Ads14550 Pseudomon	C 377	19.2	45.7	1405	11	ABZ96680	Abz96680 Human nuc
C 305	19.2	45.7	1833	11	ABD09767	Abd09767 Pseudomon	C 378	19.2	45.7	1405	11	ABD20529	Abd20529 Human pul
C 306	19.2	45.7	1980	11	ABD09903	Abd09903 Pseudomon	C 379	19.2	45.7	1608	2	AAQ25790	Aaq25790 Encodes n
C 307	19.2	45.7	2000	11	ACL36039	Acl36039 Rice stre	C 380	19.2	45.7	1664	3	AAA34862	Aaa34862 Human low
C 308	19.2	45.7	2006	8	ABT18620	Abt18620 Aspergill	C 381	19.2	45.7	1664	3	AAQ20984	Aaf20984 Human low
C 309	19.2	45.7	2070	8	ABT21034	Abt21034 Aspergill	C 382	19.2	45.7	1664	10	ABZ96678	Abz96678 Human nuc
C 310	19.2	45.7	2235	8	ABT20436	Abt20436 Aspergill	C 383	19.2	45.7	1664	10	ACA56622	Acas56622 Human sig

384	19	45.2	1664	11	ABD20527	Abd20527 Human pul	c 457	18.8	44.8	412	6	ABN23726	Abn23726 Human ORF
385	19	45.2	1664	12	AD156418	Ad156418 Human pol	458	18.8	44.8	413	10	ADD26877	Add26877 Human adi
386	19	45.2	1784	14	ABE67638	AbE67638 Rice geno	c 459	18.8	44.8	466	8	ABX41497	Abx41497 Bovine ES
387	19	45.2	1899	2	AAT05530	Aat05530 Metallopro	c 460	18.8	44.8	508	12	ACH77040	Ach77040 Human gen
388	19	45.2	1998	3	AAA34863	Aaa34863 Human ade	461	18.8	44.8	540	8	ACF72639	AcF72639 Staphyloc
389	19	45.2	1998	3	AC73842	Aac73842 Human IL-	c 462	18.8	44.8	551	5	ABA12328	Abal2328 Human ner
390	19	45.2	1998	3	AAT20985	Aat20985 Human low	c 463	18.8	44.8	563	3	AAA42386	Aaa42386 Human sec
391	19	45.2	1998	8	ABX04496	Abx04496 Human Int	c 464	18.8	44.8	597	4	AAI16648	Aai16648 Probe #65
392	19	45.2	1998	10	ABZ96679	Abz96679 Human nuc	c 465	18.8	44.8	597	4	ABA59987	AbA59987 Human foe
393	19	45.2	1998	11	ABD20528	Abd20528 Human pul	c 466	18.8	44.8	597	4	AAI139860	Aai139860 Probe #85
394	19	45.2	1998	12	ADR12173	Adr12173 Human int	c 467	18.8	44.8	597	4	ABA28396	AbA28396 Probe #68
395	19	45.2	2006	2	AAQ22978	Aaq22978 Sequence	c 468	18.8	44.8	597	4	AAK34137	Aak34137 Human bon
396	19	45.2	2006	2	AAQ22973	Aaq22973 Sequence	c 469	18.8	44.8	597	4	AAK08258	Aak08258 Human bra
397	19	45.2	2006	3	AAA34865	Aaa34865 Human ade	c 470	18.8	44.8	597	4	ABS33935	AbS33935 Human liv
398	19	45.2	2006	3	AC73823	Aac73823 Human IL-	c 471	18.8	44.8	597	6	ABS08900	AbS08900 Human gen
399	19	45.2	2006	3	AAF20987	Aaf20987 Human low	c 472	18.8	44.8	633	6	ABQ42805	AbQ42805 Oligonuc1
400	19	45.2	2006	8	ABX04477	Abx04477 Human Int	c 473	18.8	44.8	633	6	ABQ42804	AbQ42804 Oligonuc1
401	19	45.2	2006	10	ABZ96681	Abz96681 Human nuc	c 474	18.8	44.8	698	4	AAH06169	Aah06169 Human CDN
402	19	45.2	2006	11	ABD20530	Abd20530 Human pul	c 475	18.8	44.8	750	2	AAV24854	Aav24854 H. pylori
403	19	45.2	2006	12	ADJ61308	Adj61308 IL5R-X611	c 476	18.8	44.8	750	2	AAK75831	Aak75831 H. pylori
404	19	45.2	2006	12	ADO46698	Ado46698 Human oli	c 477	18.8	44.8	757	4	AAK93303	Aak93303 Human CDN
405	19	45.2	2006	12	ADR12154	Adr12154 Human int	c 478	18.8	44.8	757	4	AAK91894	Aak91894 Human CDN
406	19	45.2	2024	2	AAQ22979	Aaq22979 Sequence	c 479	18.8	44.8	757	12	ADL28321	Adl28321 5' end of
407	19	45.2	2024	2	AAQ22975	Aaq22975 Sequence	c 480	18.8	44.8	757	12	ADL29730	Adl29730 5' end of
408	19	45.2	2024	3	AA34859	Aaa34859 Human ade	c 481	18.8	44.8	763	4	AAH99124	Aah99124 Human EST
409	19	45.2	2024	3	AAF20981	Aaf20981 Human low	c 482	18.8	44.8	780	2	AAV24741	Aav24741 H. pylori
410	19	45.2	2024	10	ABZ96675	Abz96675 Human nuc	c 483	18.8	44.8	1000	14	ABE85694	AbE85694 Human pho
411	19	45.2	2024	11	ADI31989	Adi31989 Human CDN	c 484	18.8	44.8	1000	14	ABE85696	AbE85696 Human pho
412	19	45.2	2024	11	ABD20524	Abd20524 Human pul	c 485	18.8	44.8	1029	9	ACF25382	AcF25382 Rat schwa
413	19	45.2	2024	12	ADJ59303	Adj59303 Interleuk	c 486	18.8	44.8	1251	3	AAK81855	Aak81855 cDNA enco
414	19	45.2	2024	12	ADO44793	Ado44793 Human oli	c 487	18.8	44.8	1251	3	AAA08002	Aaa08002 Human Ste
415	19	45.2	2024	13	ADS84056	AdS84056 Human lym	c 488	18.8	44.8	1289	6	ABQ52169	AbQ52169 Oligonuc1
416	19	45.2	2052	2	AAT05529	Aat05529 p45 metal	c 489	18.8	44.8	1289	6	ABQ52168	AbQ52168 Oligonuc1
417	19	45.2	2052	2	AAT40133	Aat40133 Fusarium	c 490	18.8	44.8	1346	9	ADA11561	Ada11561 Human CDN
418	19	45.2	2052	2	AAV28868	Aav28868 Fusarium	c 491	18.8	44.8	1346	9	ADA11561	Ada11561 Human CDN
419	19	45.2	2052	4	ABL04747	AbL04747 Drosophil	c 492	18.8	44.8	1353	3	AAK81858	Aak81858 Ampified
420	19	45.2	2205	8	AA533176	Aa533176 Human kin	c 493	18.8	44.8	1353	3	AAA08005	Aaa08005 Human Ste
421	19	45.2	2322	6	ABA92185	AbA92185 Human NAL	c 494	18.8	44.8	1371	10	ADE56431	Ade56431 Rat gene
422	19	45.2	2322	6	ABA92206	AbA92206 Human NYX	c 495	18.8	44.8	1372	10	ADF82068	AdF82068 Leukaemia
423	19	45.2	2397	10	ADB62505	AdB62505 Human CDN	c 496	18.8	44.8	1429	10	ADK56649	AdK56649 Plant DNA
424	19	45.2	2547	5	AA575388	Aa575388 DNA enco	c 497	18.8	44.8	1495	3	ACA46957	AcA46957 Arabidops
425	19	45.2	2713	10	ADP69147	AdP69147 Human MP5	c 498	18.8	44.8	1737	5	AAC29901	Aac29901 Human h22
426	19	45.2	2876	11	ADM02841	Adm02841 Human CDN	c 499	18.8	44.8	1770	4	AAH21020	Aah21020 Bovine-de
427	19	45.2	2876	14	AEC85771	Aec85771 Human CDN	c 500	18.8	44.8	1770	13	ADT15581	AdT15581 Plant CDN
428	19	45.2	3737	2	AAV43796	Aav43796 Sequence	c 501	18.8	44.8	2000	8	ADT2758	AdT2758 Rice gene
429	19	45.2	4345	4	ABL04746	AbL04746 Staphyloc	c 502	18.8	44.8	2436	8	ABT19144	AbT19144 Aspergill
430	19	45.2	5648	4	AC89353	Aac89353 Staphyloc	c 503	18.8	44.8	2552	4	AAH16154	Aah16154 Human CDN
431	19	45.2	11355	3	AAA13486	Aaa13486 Human ade	c 504	18.8	44.8	2589	4	AAK94285	Aak94285 Human ful
432	19	45.2	11355	3	AAF20988	Aaf20988 Human low	c 505	18.8	44.8	2589	12	ADL30895	Adl30895 Full leng
433	19	45.2	11355	10	ABZ96682	Abz96682 Human nuc	c 506	18.8	44.8	2727	8	ABT18550	Abt18550 Aspergill
434	19	45.2	11355	11	ABD20531	Abd20531 Human pul	c 507	18.8	44.8	2800	14	AED04233	AeD04233 Human Hg1
435	19	45.2	28140	12	ADO47032	Ado47032 Human oli	c 508	18.8	44.8	3020	4	AAK94468	Aak94468 Human ful
436	19	45.2	39700	12	AQD97170	Aqd97170 Mouse can	c 509	18.8	44.8	3020	12	ADL31251	Adl31251 Full leng
437	19	45.2	42587	12	ADJ61642	Adj61642 Concateme	c 510	18.8	44.8	3075	14	AED73704	Aed73704 Human pla
438	19	45.2	101169	12	AQD97584	Aqd97584 Mouse can	c 511	18.8	44.8	3160	8	ACC50232	Acc50232 Breast ca
439	19	45.2	110000	14	AED76257	Continuation (4 of	c 512	18.8	44.8	3160	13	ADR25792	AdR25792 Breast ca
440	19	45.2	135090	11	ADN95683	Adn95683 Human BEC	c 513	18.8	44.8	3179	4	AAK94508	Aak94508 Human tum
441	19	45.2	210920	12	AQD97123	Aqd97123 Mouse can	c 514	18.8	44.8	3179	10	ADD29652	AdD29652 Human ful
442	18.8	44.8	80	12	ADQ95112	Adm95112 Rat antis	c 515	18.8	44.8	3179	12	ADL31329	Adl31329 Full leng
443	18.8	44.8	131	4	AAI25844	Aai25844 Probe #15	c 516	18.8	44.8	3200	2	AAK81854	Aak81854 cDNA enco
444	18.8	44.8	131	4	ABA72528	AbA72528 Human foe	c 517	18.8	44.8	3201	3	AAA08001	Aaa08001 Human Ste
445	18.8	44.8	131	4	AAI52941	Aai52941 Probe #21	c 518	18.8	44.8	3213	8	ABT20964	AbT20964 Aspergill
446	18.8	44.8	131	4	ABA38275	AbA38275 Probe #16	c 519	18.8	44.8	3263	10	ADC49713	AdC49713 Human MAS
447	18.8	44.8	131	4	AAK47105	Aak47105 Human bon	c 520	18.8	44.8	3263	13	ADP24956	AdP24956 PRO polypr
448	18.8	44.8	131	4	AAK20952	Aak20952 Human bra	c 521	18.8	44.8	3263	13	ADU05844	AdU05844 Novel bro
449	18.8	44.8	131	4	ABS46865	AbS46865 Human liv	c 522	18.8	44.8	3263	14	ADZ10045	AdZ10045 Human bre
450	18.8	44.8	131	6	ABS21334	AbS21334 Human gen	c 523	18.8	44.8	3263	15	ABE87871	Aee87871 Human CDN
451	18.8	44.8	239	4	AAH27148	Aah27148 Human lep	c 524	18.8	44.8	3268	2	AAZ40479	Aaz40479 Human STL
452	18.8	44.8	243	6	ABN25451	Abn25451 Human ORF	c 525	18.8	44.8	3283	4	AAF75337	Aaf75337 Human TGF
453	18.8	44.8	279	12	ACH90757	Ach90757 Human gen	c 526	18.8	44.8	3335	10	ABE31758	AbE31758 Human 225
454	18.8	44.8	288	14	ASC38392	Aec38392 DNA enco	c 527	18.8	44.8	3363	8	ABT20366	Abt20366 Aspergill
455	18.8	44.8	288	14	AEC38390	Aec38390 DNA enco	c 528	18.8	44.8	3403	14	ABE32284	AbE32284 Human CDN
456	18.8	44.8	393	6	ADH31558	Adh31558 Novel yea	c 529	18.8	44.8	3408	4	AAD17487	Aad17487 Human lep



C 530	18.8	44.8	3426	2	AAX59851	Aax59851 SQ ID 4	603	18.6	44.3	2271	4	ABL21710	AbL21710 Drosophil
C 531	18.8	44.8	3426	6	ABL63717	AbL63717 Breast ca	C 604	18.6	44.3	2396	4	ABL30488	AbL30488 Drosophil
C 532	18.8	44.8	3426	12	ADQ19662	AdQ19662 Human sof	C 605	18.6	44.3	2604	13	ADS58014	AdS58014 Bacterial
C 533	18.8	44.8	3426	14	AEB32239	Aeb32239 Human cDN	C 606	18.6	44.3	2615	4	ABL09765	AbL09765 Drosophil
C 534	18.8	44.8	3560	4	AAI159504	Aai159504 Human act	C 607	18.6	44.3	3054	4	ABL11272	AbL11272 Drosophil
C 535	18.8	44.8	3755	8	ABX62955	AbX62955 Human pol	C 608	18.6	44.3	3306	4	ABL13840	AbL13840 Drosophil
C 536	18.8	44.8	4130	12	ADE77047	AdE77047 Human cDN	C 609	18.6	44.3	3411	13	ADR07076	AdR07076 Full leng
C 537	18.8	44.8	4258	5	AAS78055	Aas78055 DNA encod	C 610	18.6	44.3	3478	14	ABE26903	Aeb26903 pinus rad
C 538	18.8	44.8	4622	4	ABK42254	AbK42254 Genomic s	C 611	18.6	44.3	4570	4	AAF88172	Aaf88172 A thalia
C 539	18.8	44.8	4622	9	ADB60410	AdB60410 Connectiv s	C 612	18.6	44.3	4806	4	ABL09764	AbL09764 Drosophil
C 540	18.8	44.8	4727	8	ABT117956	AbT117956 Aspergill	C 613	18.6	44.3	5082	4	ABL04626	AbL04626 Drosophil
C 541	18.8	44.8	5363	8	ABT19770	AbT19770 Aspergill	C 614	18.6	44.3	5082	4	ABL04626	AbL04626 Drosophil
C 542	18.8	44.8	6240	4	ABL06443	AbL06443 Drosophil	C 615	18.6	44.3	6375	4	ABL09610	AbL09610 Drosophil
C 543	18.8	44.8	9644	14	AEB32406	Aeb32406 Human gen	C 616	18.6	44.3	10343	4	ABL14246	AbL14246 Drosophil
C 544	18.8	44.8	9933	4	ABL06442	AbL06442 Drosophil	C 617	18.6	44.3	11006	4	AAK66281	Aak66281 Human imm
C 545	18.8	44.8	10489	4	ABL14889	AbL14889 Drosophil	C 618	18.6	44.3	24174	14	AEE04893	Aee04893 Cancer-as
C 546	18.8	44.8	11566	5	ABAI4380	AbAI4380 Human ner	C 619	18.6	44.3	28136	4	AAK69755	Aak69755 Human imm
C 547	18.8	44.8	13484	4	ABLI4888	AbLI4888 Drosophil	C 620	18.6	44.3	53267	14	ABE05141_7	AbE05141_7 Continuation (8 of
C 548	18.8	44.8	13502	5	AAH48508	Aah48508 Japanese	C 621	18.6	44.3	61510	14	ADZ13267	AdZ13267 Human can
C 549	18.8	44.8	20348	14	AEB32361	Aeb32361 Human gen	C 622	18.6	44.3	82256	12	ADO34927_3	AdO34927_3 Continuation (4 of
C 550	18.8	44.8	39746	4	ABL13398	AbL13398 Drosophil	C 623	18.6	44.3	96583	3	AAF22297_BAC	Aaf22297_BAC conta
C 551	18.8	44.8	64125	13	ABD33295	AbD33295 Human can	C 624	18.6	44.3	96593	11	ADL27146	AdL27146 Human gen
C 552	18.8	44.8	106416	4	ABLI8718	AbLI8718 Drosophil	C 625	18.6	44.3	96595	9	ADA03068	Ada03068 Human PPP
C 553	18.8	44.8	110000	2	AAX20248_05	Aax20248_05 Continuation (6 of	C 626	18.6	44.3	96595	10	ADB72806	AdB72806 Human PPP
C 554	18.8	44.8	110000	2	AAX20248_06	Aax20248_06 Continuation (7 of	C 627	18.6	44.3	96596	9	ADA66352	Ada66352 Human PPP
C 555	18.8	44.8	110000	6	ABX08336_01	AbX08336_01 Continuation (2 of	C 628	18.6	44.3	110000	13	ABD32804	AbD32804 Mouse can
C 556	18.8	44.8	110000	12	ADJ25985_01	AdJ25985_01 Continuation (2 of	C 629	18.6	44.3	127168	15	ABF75295	AbF75295 Human pol
C 557	18.8	44.8	110000	12	ADN97989_01	AdN97989_01 Continuation (2 of	C 630	18.6	44.3	127369	11	ACN44006	Acn44006 Human gen
C 558	18.8	44.8	110000	12	ADQ50281_01	AdQ50281_01 Continuation (3 of	C 631	18.6	44.3	199719	14	ABE05001	AbE05001 Cancer-as
C 559	18.8	44.8	110000	13	ABD32804_2	AbD32804_2 Continuation (3 of	C 632	18.6	44.3	215221	11	ACN44754	Acn44754 Human gen
C 560	18.8	44.8	110000	14	AEB5185_01	Aeb5185_01 Continuation (2 of	C 633	18.6	44.3	218336	8	ABQ76678	AbQ76678 Androgen
C 561	18.8	44.8	111309	2	AAX20250	Aax20250 Borrelia	C 634	18.6	44.3	236303	4	AAS11614	Aas11614 Human gen
C 562	18.8	44.8	127222	12	ADQ977301	AdQ977301 Human can	C 635	18.6	44.3	310122	13	ABD32533	AbD32533 Mouse can
C 563	18.8	44.8	152501	12	ADP67269	AdP67269 Human chr	C 636	18.6	44.3	310122	14	ADZ13032	AdZ13032 Murine ca
C 564	18.8	44.8	156296	14	ABE05181	Abe05181 Cancer-as	C 637	18.4	43.8	141	12	ADO13679	AdO13679 SNP targe
C 565	18.8	44.8	160274	14	AEB32377	Aeb32377 Human gen	C 638	18.4	43.8	227	2	AAK63450	Aak63450 Human imm
C 566	18.8	44.8	160300	14	AEB32388	Aeb32388 Human gen	C 639	18.4	43.8	300	2	AAZ14924	Aaz14924 Human gen
C 567	18.8	44.8	177851	8	AAL57272	Aal57272 ba438B23-	C 640	18.4	43.8	397	4	AAI82129	Aai82129 Human pol
C 568	18.8	44.8	262090	12	ADQ59207	AdQ59207 MSI-H car	C 641	18.4	43.8	418	14	ACL54457	AcL54457 Human col
C 569	18.8	44.8	302603	11	ADP75187	AdP75187 Human End	C 642	18.4	43.8	436	10	ABD50391	AbD50391 Primary r
C 570	18.6	44.3	23	12	ADT128532	AdT128532 Human GPC	C 643	18.4	43.8	442	14	ACL54456	AcL54456 Human col
C 571	18.6	44.3	172	8	ABT14728	AbT14728 Human Bcl	C 644	18.4	43.8	448	8	ABX43134	AbX43134 Bovine ES
C 572	18.6	44.3	221	12	ADG99151	AdG99151 Kidney di	C 645	18.4	43.8	469	4	AAK62772	Aak62772 Human imm
C 573	18.6	44.3	354	3	AAC09559	Aac09559 Human sec	C 646	18.4	43.8	484	9	ACH29613	Ach29613 Human adu
C 574	18.6	44.3	376	4	AAK60572	Aak60572 Human imm	C 647	18.4	43.8	499	14	ACL58600	AcL58600 Human col
C 575	18.6	44.3	395	9	ACH49806	Ach49806 Human leu	C 648	18.4	43.8	576	2	AAQ83892	Aaq83892 Hepatitis
C 576	18.6	44.3	411	4	AAI81107	Aai81107 Human pol	C 649	18.4	43.8	576	2	AAQ83892	Aaq83892 Hepatitis
C 577	18.6	44.3	424	9	ACH45339	Ach45339 Human foe	C 650	18.4	43.8	677	10	ADG25172	AdG25172 HCV plasm
C 578	18.6	44.3	487	9	ACH39397	Ach39397 Human foe	C 651	18.4	43.8	721	3	AAF09548	Aaf09548 Fusarium
C 579	18.6	44.3	500	9	ACH15873	Ach15873 Human adu	C 652	18.4	43.8	721	13	ADU53589	AdU53589 Fusarium
C 580	18.6	44.3	533	13	ACN56036	AcN56036 Cotton an	C 653	18.4	43.8	721	14	ADZ91592	AdZ91592 Fusarium
C 581	18.6	44.3	580	2	AAQ78105	Aaq78105 Hepatitis	C 654	18.4	43.8	751	4	AAI96348	Aai96348 Human neu
C 582	18.6	44.3	580	10	ADD333983	AdD333983 Mouse mit	C 655	18.4	43.8	830	2	AAZ16126	Aaz16126 Human gen
C 583	18.6	44.3	603	6	ABK73507	AbK73507 Bacillus	C 656	18.4	43.8	895	12	ADJ41762	AdJ41762 Plant cDN
C 584	18.6	44.3	653	13	ADS59248	AdS59248 Bacterial	C 657	18.4	43.8	957	5	AAAS90478	Aas90478 DNA encod
C 585	18.6	44.3	656	13	ADT18358	AdT18358 Plant cDN	C 658	18.4	43.8	1629	10	ADC29863	AdC29863 Fertiliza
C 586	18.6	44.3	725	13	ADR13190	AdR13190 Human can	C 659	18.4	43.8	1642	4	AAAS01169	Aas01169 Fertiliza
C 587	18.6	44.3	741	13	ADR133989	AdR133989 Cotton cD	C 660	18.4	43.8	1742	4	AAH17569	Aah17569 Human CDN
C 588	18.6	44.3	822	11	ACN888277	AcN888277 Breast ca	C 661	18.4	43.8	1749	4	AAAS01158	Aas01158 Fertiliza
C 589	18.6	44.3	934	14	AEB66366	Aeb66366 Rice geno	C 662	18.4	43.8	1749	10	ADC29841	AdC29841 Fertiliza
C 590	18.6	44.3	1040	4	AAK69263	Aak69263 Human imm	C 663	18.4	43.8	1758	4	AAK83400	Aak83400 Human imm
C 591	18.6	44.3	1194	4	ABLI3841	AbLI3841 Drosophil	C 664	18.4	43.8	1794	10	ADC29839	AdC29839 Fertiliza
C 592	18.6	44.3	1236	8	ACA54395	AcA54395 Prokaryot	C 665	18.4	43.8	1813	13	ADX54110	AdX54110 Plant ful
C 593	18.6	44.3	1494	4	AAI06935	Aai06935 Human rep	C 666	18.4	43.8	1897	13	ADU07850	AdU07850 DNA seque
C 594	18.6	44.3	1496	4	AAI03200	Aai03200 Human rep	C 667	18.4	43.8	1930	4	AAK83401	Aak83401 Human imm
C 595	18.6	44.3	1496	4	AAI06934	Aai06934 Human rep	C 668	18.4	43.8	2025	12	ADL03940	AdL03940 DNA encod
C 596	18.6	44.3	1496	4	AAI06931	Aai06931 Human rep	C 669	18.4	43.8	2223	14	AED12837	Aed12837 GH74- end
C 597	18.6	44.3	1500	3	AAZ29594	Aaz29594 Japanese	C 670	18.4	43.8	2223	14	AED12900	Aed12900 A. cellul
C 598	18.6	44.3	1642	12	ADK65901	AdK65901 Potato sa	C 671	18.4	43.8	2388	4	ABL30221	AbL30221 Drosophil
C 599	18.6	44.3	1908	13	ADS61606	AdS61606 Bacterial	C 672	18.4	43.8	2446	10	ADG63603	AdG63603 Maize sca
C 600	18.6	44.3	2046	13	ADS57707	AdS57707 Bacterial	C 673	18.4	43.8	2869	10	ADD29922	AdD29922 Acidother
C 601	18.6	44.3	2084	13	ADT45082	AdT45082 Bacterial	C 674	18.4	43.8	2869	10	ABZ77632	AbZ77632 Nucleotid
C 602	18.6	44.3					C 675	18.4	43.8	2869	12	ADO52314	AdO52314 Acidother

c 676 18.4 18.4 3197 10 ADD13037  
c 677 18.4 43.8 3688 3 AAC47167  
c 678 18.4 43.8 4283 12 ADJ38145  
c 679 18.4 43.8 4635 4 ABL30320  
c 680 18.4 43.8 4950 10 ADB53099  
c 681 18.4 43.8 4950 13 ADV41301  
c 682 18.4 43.8 6060 12 ADJ38148  
c 683 18.4 43.8 6336 6 ABL67080  
c 684 18.4 43.8 6336 6 ABLN95039  
c 685 18.4 43.8 6336 14 AED18235  
c 686 18.4 43.8 6373 2 AAV74538  
c 687 18.4 43.8 6400 12 ADJ38276  
c 688 18.4 43.8 7525 4 AAS01192  
c 689 18.4 43.8 7525 10 ADC29908  
c 690 18.4 43.8 9893 4 AAK85753  
c 691 18.4 43.8 10580 4 AAK69427  
c 692 18.4 43.8 13031 11 ADM39442  
c 693 18.4 43.8 15270 4 AAK74017  
c 694 18.4 43.8 15270 4 AAK85754  
c 695 18.4 43.8 32191 4 AAL03616  
c 696 18.4 43.8 32191 4 ABA07814  
c 697 18.4 43.8 41554 11 ACN44900  
c 698 18.4 43.8 42979 4 ABL20870  
c 699 18.4 43.8 83390 3 AAF22283  
c 700 18.4 43.8 90336 3 AAF22289  
c 701 18.4 43.8 96589 9 AAL57700  
c 702 18.4 43.8 96592 9 ADA02669  
c 703 18.4 43.8 96592 10 ADE95917  
c 704 18.4 43.8 96592 10 ADE95917  
c 705 18.4 43.8 99629 4 AAF28550  
c 706 18.4 43.8 103661 12 ADQ97712  
c 707 18.4 43.8 106344 10 ADJ79961  
c 708 18.4 43.8 110000 6 ABA901993\_3  
c 709 18.4 43.8 110000 6 ABO69245\_07  
c 710 18.4 43.8 110000 6 ABO67197\_06  
c 711 18.4 43.8 110000 6 ABO87681\_3  
c 712 18.4 43.8 110000 8 ABBX3371\_3  
c 713 18.4 43.8 110000 10 ABB81391\_2  
c 714 18.4 43.8 110000 11 ACN44932\_2  
c 715 18.4 43.8 110000 11 ACN43998\_4  
c 716 18.4 43.8 110000 12 ADN46845\_07  
c 717 18.4 43.8 110000 12 ADN47591\_13  
c 718 18.4 43.8 110000 12 ADN46123\_07  
c 719 18.4 43.8 110000 12 ADN47209\_13  
c 720 18.4 43.8 110000 12 ADN46464\_07  
c 721 18.4 43.8 110000 12 ADN47960\_13  
c 722 18.4 43.8 123192 13 ADV34995  
c 723 18.4 43.8 135462 12 ADQ97061  
c 724 18.4 43.8 139257 10 ADC89520  
c 725 18.4 43.8 189013 8 ACF62741  
c 726 18.4 43.8 189013 8 ADB20856  
c 727 18.4 43.8 189013 10 ADB87945  
c 728 18.4 43.8 189013 10 ADB96928  
c 729 18.4 43.8 189013 10 ADB92119  
c 730 18.4 43.8 190242 14 ADZ12571  
c 731 18.2 43.3 65 6 ABB51456  
c 732 18.2 43.3 122 3 AAF10806  
c 733 18.2 43.3 122 13 ADU54847  
c 734 18.2 43.3 122 14 ADU292850  
c 735 18.2 43.3 136 7 ADS68818  
c 736 18.2 43.3 178 8 ABL714715  
c 737 18.2 43.3 276 5 ABBV07935  
c 738 18.2 43.3 297 6 ABL71617  
c 739 18.2 43.3 329 4 AAS37670  
c 740 18.2 43.3 342 4 AAS51449  
c 741 18.2 43.3 342 8 ACA19445  
c 742 18.2 43.3 369 5 ABBV37958  
c 743 18.2 43.3 375 11 ABB09197  
c 744 18.2 43.3 411 4 AAL89848  
c 745 18.2 43.3 454 15 AEF78516  
c 746 18.2 43.3 457 3 AAC44528  
c 747 18.2 43.3 457 3 AAC38313  
c 748 18.2 43.3 469 3 AAC38332

749 18.2 43.3 492 4 AAK67080  
750 18.2 43.3 502 4 ABA61924  
751 18.2 43.3 502 4 AAI41846  
752 18.2 43.3 502 4 AAK36132  
753 18.2 43.3 502 4 AAK10233  
754 18.2 43.3 502 4 ABS35825  
755 18.2 43.3 502 6 ABS10268  
756 18.2 43.3 508 14 ABA36476  
757 18.2 43.3 566 10 ADRK58514  
758 18.2 43.3 647 13 ADR65375  
759 18.2 43.3 647 13 ADT16840  
760 18.2 43.3 704 13 ADX47448  
761 18.2 43.3 743 4 AAH05826  
762 18.2 43.3 769 13 ADO83351  
763 18.2 43.3 776 4 AAI95091  
764 18.2 43.3 801 3 AAC46622  
765 18.2 43.3 811 4 AAI94653  
766 18.2 43.3 811 8 ACA56923  
767 18.2 43.3 811 13 ADO82213  
768 18.2 43.3 831 8 ACA34548  
769 18.2 43.3 983 13 ADS45289  
770 18.2 43.3 1008 6 ABBN67751  
771 18.2 43.3 1010 2 AAX13969  
772 18.2 43.3 1011 13 ADV85016  
773 18.2 43.3 1272 4 AAF54079  
774 18.2 43.3 1272 4 AAF54080  
775 18.2 43.3 1273 4 AAF54074  
776 18.2 43.3 1273 4 AAF54074  
777 18.2 43.3 1273 4 AAF54075  
778 18.2 43.3 1273 4 AAF54076  
779 18.2 43.3 1273 4 AAF54073  
780 18.2 43.3 1276 4 AAF54077  
781 18.2 43.3 1276 4 AAF54078  
782 18.2 43.3 1400 14 AEE14143  
783 18.2 43.3 1427 5 AAC82742  
784 18.2 43.3 1427 5 AAC82743  
785 18.2 43.3 1427 6 AAK99774  
c 786 18.2 43.3 1493 12 ADO25994  
c 787 18.2 43.3 1496 10 ADB53828  
c 788 18.2 43.3 1496 13 ADV41755  
789 18.2 43.3 1707 6 AAI71008  
790 18.2 43.3 1814 13 ADX09014  
791 18.2 43.3 1814 14 AEB70441  
792 18.2 43.3 1893 14 ADZ62955  
793 18.2 43.3 1903 13 ADB60050  
c 794 18.2 43.3 1910 11 ADM02479  
c 795 18.2 43.3 1910 14 AEC85409  
c 796 18.2 43.3 1960 4 ABL06493  
c 797 18.2 43.3 1971 4 ABA55251  
798 18.2 43.3 1971 4 ABS28575  
799 18.2 43.3 2000 8 ADA72509  
c 800 18.2 43.3 2000 8 ADA72571  
c 801 18.2 43.3 2151 10 ACF69072  
c 802 18.2 43.3 2235 4 AAI92780  
c 803 18.2 43.3 2283 5 AAS73030  
c 804 18.2 43.3 2304 8 ACA34716  
c 805 18.2 43.3 2329 5 AAS90866  
806 18.2 43.3 2352 2 AAX14325  
807 18.2 43.3 2458 14 AEE10501  
808 18.2 43.3 2481 4 AAI14242  
809 18.2 43.3 2728 13 ADQ38340  
810 18.2 43.3 2758 13 ACN43105  
811 18.2 43.3 2766 13 ACN43104  
812 18.2 43.3 2771 13 ACN43838  
813 18.2 43.3 2775 3 AAA54034  
814 18.2 43.3 2777 13 ADQ38339  
815 18.2 43.3 2781 1 AAN40177  
816 18.2 43.3 2781 1 AAN40142  
817 18.2 43.3 2792 2 AAX28626  
818 18.2 43.3 2802 8 AAT02460  
819 18.2 43.3 2804 8 AAV77053  
820 18.2 43.3 2831 13 ADQ38341  
821 18.2 43.3 2880 8 ACC00311

AAK67080 Human imm  
ABa61924 Human foe  
AAI41846 Probe #10  
AAK36132 Human bon  
AAK10233 Human bra  
ABa35825 Human liv  
ABs10268 Human gen  
AaA36476 Tomato th  
ADK58514 Plant DNA  
ADr65375 Cotton cD  
ADT16840 Plant cDN  
ADx47448 Plant ful  
AHO5826 Human cDN  
ADO83351 Plant ful  
AAI95091 Human neu  
AAC46622 Zea mays  
AAI94653 Human neu  
ACA56923 Human adi  
ADO82213 Plant ful  
ACA34548 Prokaryot  
ADa5289 Bacterial  
ABn67751 Streptoco  
AAX13969 H. pylori  
ADV85016 Streptoco  
AAF54079 hFIX gene  
AAF54080 hFIX gene  
AAF54074 hFIX gene  
AAF54075 hFIX gene  
AAF54076 hFIX gene  
AAF54073 hFIX gene  
AAF54077 hFIX gene  
AAF54078 hFIX gene  
AeE14143 Transgene  
AC82742 C. sinens  
AC82743 C. sinens  
AAK99774 1427nt DN  
ADO25994 Barley ch  
AD53828 Primary r  
ADV1755 Rat cardr  
AAI71008 Human Fac  
ADX09014 Plant ful  
Aeb70441 Insulin-1  
ADZ62955 Murine C4  
AD60050 Plant ful  
ADM02479 Human cDN  
AEC85409 Human cDN  
ABl06493 Drosophil  
ABA55251 Human foe  
ABs28575 Human liv  
ADA72509 Rice gene  
ADA72571 Rice gene  
ACF69072 Photorhab  
AAI92780 Human pol  
AAS73030 DNA encod  
ACA34716 Prokaryot  
AAS90866 DNA encod  
AAX14325 H. pylori  
AeE10501 Human rec  
AAH14242 Human cDN  
ADQ38340 Human SNP  
ACN43105 Human dia  
ACN43104 Human dia  
ADQ38338 Human SNP  
AAa54034 Human fac  
ADQ38339 Human SNP  
AAN40177 Part of t  
AAN40142 Part of t  
AAX28626 Nuclcotid  
AAT02460 Human Fac  
AAV77053 Nuclcotid  
ADQ38341 Human SNP  
ACC00311 Halohydril

C 822	18.2	43.3	2905	10	ABT411705	Abt411705 Toxicity	895	18.2	43.3	128978	6	ABK83459	Abk83459 Human cdn
C 823	18.2	43.3	2994	5	AAH78232	Aah78232 Nucleotid	896	18.2	43.3	128978	8	AAD54587	Aad54587 Human LIM
C 824	18.2	43.3	3168	6	ABN79811	Abn79811 Fungal 2B	897	18.2	43.3	128978	13	AD52994	Ad52994 Drug ther
C 825	18.2	43.3	3187	12	ADO17407	Ado17407 Human sof	898	18.2	43.3	128978	14	ADX07226	Adx07226 Cyclin-de
C 826	18.2	43.3	3306	3	AAA87701	Aaa87701 Human sec	C 899	18.2	43.3	130414	11	ACN44512	Acn44512 Human CA
C 827	18.2	43.3	3306	8	ADA40549	Ada40549 Human sec	C 900	18.2	43.3	143412	11	ACN44512	Acn44512 Human gen
C 828	18.2	43.3	3306	8	ACC50852	Acc50852 Human sec	C 901	18.2	43.3	155462	12	ADQ97394	Adq97394 Mouse can
C 829	18.2	43.3	3306	8	ABZ71477	Abz71477 Secreted	C 902	18.2	43.3	158001	12	ADL17884	Adl17884 Human pho
C 830	18.2	43.3	3306	9	ADB91444	Adb91444 Human sec	C 903	18.2	43.3	161652	12	ADP13118	Adp13118 Hypermech
C 831	18.2	43.3	3306	10	ADC73384	Adc73384 Human sec	C 904	18.2	43.3	161652	12	ADI37264	Adi37264 Hypermech
C 832	18.2	43.3	3306	10	ADA56695	Ada56695 Gene enco	C 905	18.2	43.3	170834	10	AAD62833	Aad62833 Human BAC
C 833	18.2	43.3	3308	3	AAA87681	Aaa87681 Human sec	C 906	18.2	43.3	171936	6	ABS56565	Abs56565 Human SUL
C 834	18.2	43.3	3308	8	ADA40228	Ada40228 Human sec	C 907	18.2	43.3	171936	12	ADN16205	Adn16205 Mouse sul
C 835	18.2	43.3	3308	8	ACC50671	Acc50671 Human sec	C 908	18.2	43.3	171936	10	ADL13775	Adl13775 Osteoarth
C 836	18.2	43.3	3308	8	ABZ71365	Abz71365 Secreted	C 909	18.2	43.3	173805	10	ABD133462	Abd133462 Human can
C 837	18.2	43.3	3308	9	ADB91306	Adb91306 Human sec	C 910	18.2	43.3	196063	13	ABD33462	Abd33462 Human can
C 838	18.2	43.3	3308	9	ADC73354	Adc73354 Human sec	C 911	18.2	43.3	222930	6	ABK84349	Abk84349 Human cdn
C 839	18.2	43.3	3308	10	ADA56388	Ada56388 Gene enco	C 912	18.2	43.3	222930	14	AED17977	Aed17977 Fibrotic
C 840	18.2	43.3	3397	5	ABV25291	Abv25291 Human pro	C 913	18.2	43.3	243072	10	ACF65382	Acf65382 Photorhab
C 841	18.2	43.3	3799	12	ADQ22263	Adq22263 Human sof	C 914	18.2	43.3	252907	13	ABD33694	Abd33694 Human can
C 842	18.2	43.3	3832	6	ABN59890	Abn59890 Novel hum	C 915	18.2	43.3	252907	13	ABD33020	Abd33020 Mouse can
C 843	18.2	43.3	4187	4	ABL29892	AbL29892 Drosophil	C 916	18.2	43.3	259500	15	AEF07224	Aef07224 PFTK1 loc
C 844	18.2	43.3	4268	4	ABL06492	AbL06492 Drosophil	C 917	18.2	43.3	288563	14	ASE05135	Ase05135 Cancer-as
C 845	18.2	43.3	4670	12	ADO35651	Ado35651 Novel mou	C 918	18.2	43.3	349980	6	ABQ81844	Abq81844 Bifidobac
C 846	18.2	43.3	7260	6	ABK35504	Abk35504 Human end	C 919	18.2	42.9	142	2	AAV79046	Aav79046 Staphyloc
C 847	18.2	43.3	7260	6	ABK35561	Abk35561 Gene IGF1	C 920	18.2	42.9	146	8	ABT14725	Abt14725 Human Bcl
C 848	18.2	43.3	7260	6	ABT11091	Abt11091 Human bre	C 921	18.2	42.9	377	13	ADR63349	Adr63349 Cotton cD
C 849	18.2	43.3	7260	6	ABK84583	Abk84583 Human cdn	C 922	18.2	42.9	395	4	AAI85217	Aai85217 Human pol
C 850	18.2	43.3	7260	6	ABK64812	Abk64812 Human ben	C 923	18.2	42.9	400	2	AAV78280	Aav78280 Staphyloc
C 851	18.2	43.3	7260	6	ABN97244	Abn97244 Gene #374	C 924	18.2	42.9	423	8	ACF73809	Acf73809 Staphyloc
C 852	18.2	43.3	7260	9	ADA26452	Ada26452 Human ins	C 925	18.2	42.9	453	8	ABX37646	Abx37646 Bovine ES
C 853	18.2	43.3	7260	10	ADG33110	Adg33110 Human DNA	C 926	18.2	42.9	460	12	ADQ04184	Adq04184 Maize tra
C 854	18.2	43.3	7260	11	ADP65015	Adp65015 Human ins	C 927	18.2	42.9	519	14	ADY65571	Ady65571 S. mangon
C 855	18.2	43.3	7260	14	ADV16941	Adv16941 Human ins	C 928	18.2	42.9	527	12	ACH80203	Ach80203 Human gen
C 856	18.2	43.3	7260	14	ADY16236	Ady16236 DNA enco	C 929	18.2	42.9	535	4	AAF25733	Aaf25733 C. purpur
C 857	18.2	43.3	7260	14	ADY16234	Ady16234 DNA enco	C 930	18.2	42.9	613	14	ADZ61968	Adz61968 Murine Mg
C 858	18.2	43.3	7260	14	ABE96741	Abe96741 Human IGF	C 931	18.2	42.9	636	2	AAV77975	Aav77975 Staphyloc
C 859	18.2	43.3	7260	15	ASF87543	Aef87543 Human liv	C 932	18.2	42.9	685	14	ADY71697	Ady71697 Coryneb
C 860	18.2	43.3	7476	13	ABD32850	Abd32850 Human can	C 933	18.2	42.9	703	9	ADB82424	Adb82424 Human cdn
C 861	18.2	43.3	7911	14	ABE70428	Abe70428 Insulin-1	C 934	18.2	42.9	732	6	ABQ89481	Abq89481 Human pro
C 862	18.2	43.3	7935	14	ABE70427	Abe70427 Insulin-1	C 935	18.2	42.9	736	4	ABQ89481	Abq89481 Human pro
C 863	18.2	43.3	11022	4	AA341739	Aas41739 Genomic s	C 936	18.2	42.9	768	4	ABL14975	AbL14975 Drosophil
C 864	18.2	43.3	11022	4	AAK80617	Aak80617 Human imm	C 937	18.2	42.9	771	11	ADP19671	Adp19671 Leptospir
C 865	18.2	43.3	11693	6	ABN85367	Abn85367 Spumaret	C 938	18.2	42.9	768	4	ABL14975	AbL14975 Drosophil
C 866	18.2	43.3	11757	6	ABN85366	Abn85366 Spumaret	C 939	18.2	42.9	777	3	AAK47062	Aak47062 Arabidops
C 867	18.2	43.3	14518	6	AB579053	Ab579053 E. coli C	C 940	18.2	42.9	821	4	AAK75860	Aak75860 Human imm
C 868	18.2	43.3	14518	10	ADH80620	Adh80620 Escherich	C 941	18.2	42.9	852	3	AAA13883	Aaa13883 L. interc
C 869	18.2	43.3	15153	12	ADG93315	Adg93315 DEN3 (Sle	C 942	18.2	42.9	866	3	AAK54853	Aak54853 Arabidops
C 870	18.2	43.3	16291	4	AA3411741	Aas411741 Genomic s	C 943	18.2	42.9	868	3	AAK38487	Aak38487 Arabidops
C 871	18.2	43.3	16291	4	AAK80619	Aak80619 Human imm	C 944	18.2	42.9	877	3	AAK41927	Aak41927 Arabidops
C 872	18.2	43.3	17929	10	ADC86000	Adc86000 Human GPC	C 945	18.2	42.9	878	14	ABE89417	Abe89417 Isolated
C 873	18.2	43.3	20776	4	ABL30040	AbL30040 Drosophil	C 946	18.2	42.9	896	13	ADR63326	Adr63326 Cotton CD
C 874	18.2	43.3	24333	12	ADU12497	AdJ12497 DNA fragm	C 947	18.2	42.9	899	11	ACN85147	Acn85147 Breast ca
C 875	18.2	43.3	24333	12	ADU12606	AdJ12606 DNA fragm	C 948	18.2	42.9	919	3	ACN51838	Acn51838 Arabidops
C 876	18.2	43.3	38059	4	AAF54018	Aaf54018 Human fac	C 949	18.2	42.9	940	6	ABQ88077	Abq88077 Human GPC
C 877	18.2	43.3	38059	6	ABN95627	Abn95627 Gene #212	C 950	18.2	42.9	966	5	ABQ88077	Abq88077 Human GPC
C 878	18.2	43.3	38059	10	ADD71098	AdD71098 Human coa	C 951	18.2	42.9	1041	3	AAK39294	Aak39294 Arabidops
C 879	18.2	43.3	41104	6	ADJ36260	AdJ36260 Human G-p	C 952	18.2	42.9	1041	4	AAK39294	Aak39294 Arabidops
C 880	18.2	43.3	54993	11	ACN45086	Acn45086 Human gen	C 953	18.2	42.9	1041	8	ACA34847	Aca34847 Prokaryot
C 881	18.2	43.3	59560	4	AAK80620	Aak80620 Human imm	C 954	18.2	42.9	1042	3	AAK54852	Aak54852 Arabidops
C 882	18.2	43.3	59560	4	AAK71550	Aak71550 Human imm	C 955	18.2	42.9	1080	8	ACA39731	AcA39731 Prokaryot
C 883	18.2	43.3	68193	12	ADQ97400	Adq97400 Mouse can	C 956	18.2	42.9	1097	2	AAK14425	Aax14425 H. pylori
C 884	18.2	43.3	73507	11	ACN44612	Acn44612 Mouse gen	C 957	18.2	42.9	1129	11	ACN85653	Acn85653 Breast ca
C 885	18.2	43.3	95596	13	ADV87741	Adv87741 Streptoco	C 958	18.2	42.9	1137	4	AAF72044	Aaf72044 Coryneb
C 886	18.2	43.3	95596	13	ADV78994	Adv78994 Streptoco	C 959	18.2	42.9	1200	13	ADT42031	Adt42031 Bacterial
C 887	18.2	43.3	104900	13	ABD32848	Abd32848 Human can	C 960	18.2	42.9	1236	3	AAK36186	Aak36186 Arabidops
C 888	18.2	43.3	110000	4	AAF22303_4	Continuation (5 of	C 961	18.2	42.9	1275	13	ADS60929	AdS60929 Bacterial
C 889	18.2	43.3	110000	4	AAI99682_17	Continuation (18 o	C 962	18.2	42.9	1373	13	ADT15055	Adt15055 Plant CDN
C 890	18.2	43.3	110000	4	AAI99683_17	Continuation (18 o	C 963	18.2	42.9	1419	15	AEF00581	Aef00581 Human des
C 891	18.2	43.3	110000	6	AEF71527_03	Continuation (4 of	C 964	18.2	42.9	1533	15	AEF00579	Aef00579 Human des
C 892	18.2	43.3	110000	10	ACF67367_19	Continuation (20 o	C 965	18.2	42.9	1596	15	AEF00577	Aef00577 Human des
C 893	18.2	43.3	110000	13	ADV81204_03	Continuation (4 of	C 966	18.2	42.9	1902	4	ABL05301	AbL05301 Drosophil
C 894	18.2	43.3	121151	15	ABF74657	Aef74657 Human pol	C 967	18.2	42.9	2028	14	AED20066	Aed20066 Human ova

c 968 18 42.9 2094 4 ABL20545  
c 969 18 42.9 2120 6 ABL20545  
c 970 18 42.9 2120 6 ABL20545  
c 971 18 42.9 2120 6 ABL20545  
c 972 18 42.9 2145 13 ADH80608  
c 973 18 42.9 2160 13 ADH80608  
c 974 18 42.9 2165 13 ADH80608  
c 975 18 42.9 2232 13 ADH80608  
c 976 18 42.9 2293 1 AAN50491  
c 977 18 42.9 2297 1 AAN50491  
c 978 18 42.9 2320 13 ADH80608  
c 979 18 42.9 2323 2 AAV22667  
c 980 18 42.9 2316 9 ADA48505  
c 981 18 42.9 2725 6 AAL70709  
c 982 18 42.9 2760 13 ACN38980  
c 983 18 42.9 2760 14 ADZ26051  
c 984 18 42.9 3015 13 ADX29049  
c 985 18 42.9 3108 10 ACF36191  
c 986 18 42.9 3108 13 ADQ95764  
c 987 18 42.9 3108 14 ADZ71329  
c 988 18 42.9 3237 5 AAS87056  
c 989 18 42.9 3902 4 ABL05300  
c 990 18 42.9 4717 13 ADR08268  
c 991 18 42.9 4783 13 ADR07170  
c 992 18 42.9 5195 6 ABL33596  
c 993 18 42.9 6507 8 ADA89838  
c 994 18 42.9 6520 12 ADO79327  
c 995 18 42.9 6520 12 ADO23329  
c 996 18 42.9 7100 4 ABL29168  
c 997 18 42.9 7430 2 AAV74571  
c 998 18 42.9 10325 4 AAK67805  
c 999 18 42.9 10710 6 ABL79997  
1000 18 42.9 10818 12 ADO07431

## ALIGNMENTS

RESULT 1  
ADT96418  
ID ADT96418 standard; cDNA; 497 BP.  
XX AC ADT96418;  
XX DT 16-DEC-2004 (first entry)  
XX DE Colon cancer associated human cDNA sequence #1925.  
XX KW Colon cancer; T cell; tumour protein; C634S; C635S; C637S; C640S; C636S;  
KW humoral immune response; cellular immune response; cytostatic;  
KW immunostimulant; human; ss.  
XX OS Homo sapiens.  
XX PN US2003087818-A1.  
XX DT 08-MAY-2003.  
XX PF 01-FEB-2002; 2002US-00066543.  
XX PR 02-FEB-2001; 2001US-0267400P.  
XX PR 07-FEB-2001; 2001US-0267382P.  
XX PR 11-MAY-2001; 2001US-0290322P.  
XX PR 12-JUL-2001; 2001US-0305265P.  
XX PR 16-AUG-2001; 2001US-0313077P.  
XX PA (CORI-) CORIXA CORP.  
XX PI Jiang Y, Chenault RA, Xu J, Indirias CY, Lodes MJ, Secretist H;  
PI Carter D, Fanger GR, Smith CL, Durham M, Stolk JA;  
XX WPI; 2003-040540/03.

PT New isolated nucleic acids and polypeptides capable of eliciting a  
PT humoral and/or cellular immune response, useful for diagnosing,  
XX preventing or treating cancer, particularly colon cancer.  
PS Claim 1; SEQ ID NO 1937; 87pp; English.  
XX The invention relates to polynucleotide and polypeptide sequences  
CC associated with cancer, particularly colon cancer. Also disclosed are (i)  
CC an expression vector comprising the polynucleotide, (ii) a host cell  
CC transformed or transfected with the expression vector, (iii) an isolated  
CC antibody, or its antigen-binding fragment, which specifically binds to  
CC the polypeptide, (iv) a method of detecting or determining the presence  
CC of cancer in a patient, (v) a fusion protein comprising at least one of  
CC the polypeptides, (vi) an oligonucleotide that hybridises to the  
CC polynucleotide sequence under highly stringent conditions, and (vii) a  
CC method of stimulating and/or expanding T cells specific for a tumour  
CC protein. The polypeptide specifically comprises the amino acid sequence  
CC of C634S, C635S, C637S, C640S, C636S or one of the potential open reading  
CC frames (ORFs) of C636S. These polypeptides are encoded by the  
CC polynucleotide sequences, where both are capable of eliciting a humoral  
CC and/or cellular immune response. The polynucleotides, polypeptides, and  
CC antibodies are useful for diagnosing, preventing or treating cancer,  
CC particularly colon cancer. The polynucleotide and polypeptide sequences  
CC are also useful in DNA strand invasion, antisense inhibition, mutational  
CC analysis, nucleic acid purification, isolation of transcriptionally  
CC active genes, blocking or transcription factor binding, genome cleavage  
CC or in situ hybridisation, and as enhancers of transcription or  
CC biomarkers. This sequence represents a human colon cancer associated  
CC cDNA. Note: The sequence data for this patent was obtained in electronic  
CC format directly from the USPTO web site at seqdata.uspto.gov  
XX Sequence 497 BP; 126 A; 130 C; 129 G; 112 T; 0 U; 0 Other;  
SQ Query Match 100.0%; Score 42; DB 11; Length 497;  
Best Local Similarity 100.0%; Pred. No. 1.2e-07;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GCCACGGTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
Db 344 GCCACGGTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 385  
RESULT 2  
ADX42900  
ID ADX42900 standard; cDNA; 497 BP.  
XX AC ADX42900;  
XX DT 21-APR-2005 (first entry)  
XX DE Human cDNA encoding colon cancer protein SEQ ID NO 1937.  
XX KW Cytostatic; Immunostimulant; therapy; diagnosis; colon cancer; neoplasm;  
KW ss; gene.  
XX OS Homo sapiens.  
XX PN WO200274156-A2.  
XX PD 26-SEP-2002.  
XX PF 01-FEB-2002; 2002WO-US002870.  
XX PR 02-FEB-2001; 2001US-0267400P.  
XX PR 07-FEB-2001; 2001US-0267382P.  
XX PR 11-MAY-2001; 2001US-0290322P.  
XX PR 12-JUL-2001; 2001US-0305265P.  
XX PR 16-AUG-2001; 2001US-0313077P.  
XX PA (CORI-) CORIXA CORP.  
XX PI Jiang Y, Chenault RA, Xu J, Indirias CY, Lodes MJ, Secretist H;  
PI Carter D, Fanger GR, Smith CL, Durham M, Stolk JA;

```
XX WPI; 2003-040540/03.
DR
XX New isolated nucleic acids and polypeptides capable of eliciting a
PT humoral and/or cellular immune response, useful for diagnosing,
PT preventing or treating cancer, particularly colon cancer.
PS
XX Claim 1; SEQ ID NO 1937; 244pp; English.
XX
CC The invention relates to a new isolated nucleic acid. The nucleic acids,
CC polypeptides, antibodies are useful for diagnosing, preventing or
CC treating cancer, particularly colon cancer. The nucleic acid and
CC polypeptides are also useful in DNA strand invasion, antisense
CC inhibition, mutational analysis, nucleic acid purification, isolation of
CC transcriptionally active genes, blocking or transcription factor binding,
CC genome cleavage or in situ hybridization, and as enhancers of
CC transcription or biomarkers. The kits are useful for detecting antibody
CC binding. The present sequence represents a human cDNA encoding a colon
CC cancer protein.
XX
SQ Sequence 497 BP; 126 A; 130 C; 129 G; 112 T; 0 U; 0 Other;
Query Match 100.0%; Score 42; DB 11; Length 497;
Best Local Similarity 100.0%; Pred. No. 1.2e-07;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42
DB 344 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 385
RESULT 3
ADT96292/c
ID ADT96292 standard; cDNA; 552 BP.
XX
AC ADT96292;
XX
DT 16-DEC-2004 (first entry)
XX
DE Colon cancer associated human cDNA sequence #1799.
XX
KW Colon cancer; T cell; tumour protein; C634S; C637S; C640S; C636S;
KW humoral immune response; cellular immune response; cytostatic;
KW immunostimulant; human; ss.
XX
OS Homo sapiens.
XX
PN US2003087818-A1.
XX
PD 08-MAY-2003.
XX
PF 01-FEB-2002; 2002US-00066543.
XX
PR 02-FEB-2001; 2001US-0267400P.
PR 07-FEB-2001; 2001US-0267382P.
PR 11-MAY-2001; 2001US-0290322P.
PR 12-JUL-2001; 2001US-0305265P.
PR 16-AUG-2001; 2001US-0313077P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Jiang Y, Chenault RA, Xu J, Indirias CV, Lodes MJ, Secrlist H;
PI Carter D, Fanger GR, Smith CL, Durham M, Stolk JA;
XX
WPI; 2003-040540/03.
XX
DR New isolated nucleic acids and polypeptides capable of eliciting a
PT humoral and/or cellular immune response, useful for diagnosing,
PT preventing or treating cancer, particularly colon cancer.
XX
PS Claim 1; SEQ ID NO 1811; 87pp; English.
XX
XX The invention relates to polynucleotide and polypeptide sequences
CC associated with cancer, particularly colon cancer. Also disclosed are (i)
CC an expression vector comprising the polynucleotide, (ii) a host cell
CC transformed or transfected with the expression vector, (iii) an isolated
CC antibody, or its antigen-binding fragment, which specifically binds to
CC the polypeptide, (iv) a method of detecting or determining the presence
CC of cancer in a patient, (v) a fusion protein comprising at least one of
CC the polypeptides, (vi) an oligonucleotide that hybridises to the
CC polynucleotide sequence under highly stringent conditions, and (vii) a
CC method of stimulating and/or expanding T cells specific for a tumour
CC protein. The polypeptide specifically comprises the amino acid sequence
CC of C634S, C635S, C637S, C640S, C636S or one of the potential open reading
CC frames (ORFs) of C636S. These polypeptides are encoded by the
CC polynucleotide sequences, where both are capable of eliciting a humoral
CC and/or cellular immune response. The polynucleotides, polypeptides, and
CC antibodies are useful for diagnosing, preventing or treating cancer,
CC particularly colon cancer. The polynucleotide and polypeptide sequences
CC are also useful in DNA strand invasion, antisense inhibition, mutational
CC analysis, nucleic acid purification, isolation of transcriptionally
CC active genes, blocking or transcription factor binding, genome cleavage
CC or in situ hybridisation, and as enhancers of transcription or
CC biomarkers. This sequence represents a human colon cancer associated
CC cDNA. Note: The sequence data for this patent was obtained in electronic
CC format directly from the USPTO web site at seqdata.uspto.gov
XX
SQ Sequence 552 BP; 126 A; 141 C; 148 G; 137 T; 0 U; 0 Other;
Query Match 100.0%; Score 42; DB 11; Length 552;
Best Local Similarity 100.0%; Pred. No. 1.2e-07;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42
DB 154 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 113
RESULT 4
ADX42774/c
ID ADX42774 standard; cDNA; 552 BP.
XX
AC ADX42774;
XX
DT 21-APR-2005 (first entry)
XX
DE Human cDNA encoding colon cancer protein SEQ ID NO 1811.
XX
KW Cytostatic; Immunostimulant; therapy; diagnosis; colon cancer; neoplasm;
KW ss; gene.
XX
OS Homo sapiens.
XX
PN WO200274156-A2.
XX
PD 26-SEP-2002.
XX
PR 01-FEB-2002; 2002WO-US002870.
XX
PR 02-FEB-2001; 2001US-0267400P.
PR 07-FEB-2001; 2001US-0267382P.
PR 11-MAY-2001; 2001US-0290322P.
PR 12-JUL-2001; 2001US-0305265P.
PR 16-AUG-2001; 2001US-0313077P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Jiang Y, Chenault RA, Xu J, Indirias CV, Lodes MJ, Secrlist H;
PI Carter D, Fanger GR, Smith CL, Durham M, Stolk JA;
XX
WPI; 2003-040540/03.
XX
DR New isolated nucleic acids and polypeptides capable of eliciting a
PT humoral and/or cellular immune response, useful for diagnosing,
PT preventing or treating cancer, particularly colon cancer.
XX
PS Claim 1; SEQ ID NO 1811; 87pp; English.
XX
XX The invention relates to polynucleotide and polypeptide sequences
```

```
PS Claim 1; SEQ ID NO 1811; 244pp; English.
XX
CC The invention relates to a new isolated nucleic acid. The nucleic acids,
CC polypeptides, antibodies are useful for diagnosing, preventing or
CC treating cancer, particularly colon cancer. The nucleic acid and
CC polypeptides are also useful in DNA strand invasion, antisense
CC inhibition, mutational analysis, nucleic acid purification, isolation of
CC transcriptionally active genes, blocking or transcription factor binding,
CC genome cleavage or in situ hybridization, and as enhancers of
CC transcription or biomarkers. The kits are useful for detecting antibody
CC binding. The present sequence represents a human cDNA encoding a colon
CC cancer protein.
XX
SQ Sequence 552 BP; 126 A; 141 C; 148 G; 137 T; 0 U; 0 Other;
Query Match 100.0%; Score 42; DB 11; Length 552;
Best Local Similarity 100.0%; Pred. No. 1.2e-07;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCCCAGCGTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42
Db 154 GCCCAGCGTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 113
RESULT 5
AAZ90046
ID AAZ90046 standard; cDNA; 603 BP.
AC AAZ90046;
XX
DT 09-MAY-2000 (first entry)
XX
DE Hydrophobic domain containing protein clone HP10549 coding sequence.
KW Hydrophobic domain; clone HP10549; nutritional supplement; SCID; HIV;
KW cell proliferation; immune stimulant; immune deficiency; tumour; pain;
KW rheumatoid arthritis; insulin dependent diabetes mellitus; fertility;
KW myasthenia gravis; haematopoiesis regulator; tissue growth; depression;
KW anti-inflammatory; infection; bodily characteristic; ss.
XX
OS Homo sapiens.
XX
PN WO200000506-A2.
XX
PD 06-JAN-2000.
XX
PF 18-JUN-1999; 99WO-JP003242.
XX
PR 26-JUN-1998; 98JP-00180008.
XX
PA (SAGA ) SAGAMI CHEM RES CENT.
PA (PROT-) PROTEGENE INC.
PI Kato S, Kimura T;
XX
WPI; 2000-160665/14.
DR P-PSDB; AAY78809.
XX
PT Novel human proteins having hydrophobic domains used for research and
PT diagnostic purposes.
XX
PS Claim 3; Page 88; 117pp; English.
XX
CC This sequence represents the hydrophobic domain containing protein, clone
CC HP10549 coding region. The sequence is isolated from a human stomach
CC cancer cell line. The invention relates to human proteins with
CC hydrophobic domains, the DNA and the cDNA encoding them. The
CC polynucleotides and proteins are predicted to have biological activities
CC which make them suitable for treating, preventing or ameliorating medical
CC conditions in humans and animals. Suggested activities include
CC nutritional activity (nutritional source or supplement); cytokine and
CC cell proliferation/differentiation activity; immune stimulating (e.g. as
CC vaccines) or suppressing activity (e.g. to treat various immune
CC
CC deficiencies such as SCIDS or HIV, connective tissue disease, systemic
CC lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary
CC inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin
CC dependent diabetes mellitus, myasthenia gravis, graft-versus-host disease
CC and autoimmune inflammatory eye disease, as well as asthma, allergies and
CC organ transplantation); haematopoiesis regulating activity (e.g. in
CC treatment of myeloid or lymphoid cell deficiencies); tissue growth
CC activity (e.g. wound healing and tissue repair, ulcers, burns,
CC periodontal disease); activin/inhibin activity; chemotactic/chemokinetic
CC activity; haemostatic and thrombolytic activity (e.g. treating
CC haemophilias); receptor/ligand activity; anti-inflammatory activity; and
CC tumour inhibition activity. The polynucleotides are also stated to be
CC useful for gene therapy. Other activities include inhibiting infections
CC caused by bacteria, fungi, viruses and other parasites (e.g. Hepatitis,
CC malaria); effecting bodily characteristics such as, e.g. weight, colour,
CC skin, effecting biorhythms or cardiac cycles; enhancing fertility;
CC treatment of depression; treatment of pain; hormonal or endocrine
CC activity. The polynucleotides may also be used for recombinant expression
CC of the protein
XX
SQ Sequence 603 BP; 133 A; 179 C; 137 G; 154 T; 0 U; 0 Other;
Query Match 100.0%; Score 42; DB 3; Length 603;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCCCAGCGTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42
Db 550 GCCCAGCGTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 591
RESULT 6
ABQ58527/c
ID ABQ58527 standard; cDNA; 620 BP.
XX
AC ABQ58527;
XX
DT 02-AUG-2002 (first entry)
XX
DE Human colon cancer related nucleotide sequence SEQ ID NO:2222.
XX
KW Human; colon cancer; cancer; tissue profiling; forensic; mapping;
KW genetic analysis; diagnostic; antisense therapy; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200229086-A2.
XX
PD 11-APR-2002.
XX
PF 02-OCT-2001; 2001WO-US030732.
XX
PR 02-OCT-2000; 2000US-0237271P.
XX
PA (FARB ) BAYER CORP.
XX
PI Burgess C, Astle JH, Carroll E, Catino TJ, Dwivedi P, Molino GA;
PI Thiagalingam A, Lewis ME;
XX
WPI; 2002-426115/45.
XX
PT New isolated nucleic acid that is differentially expressed in cancer
PT tissues useful for determining the presence of colon cancer in a cell or
PT tissue type, and in antisense therapy.
XX
PS Claim 1; Fig 1; 796pp; English.
XX
CC ABQ56306 to ABQ60787 represent isolated nucleic acids (I) differentially
CC expressed in cancer tissues. ABB78993 to ABB79004 represent proteins
CC encoded by the ABQ60776 to ABQ60787 nucleic acid sequences. (I) can be
CC used in antisense therapy. An antibody immunoreactive with a polypeptide
CC encoded by (I) is useful for detecting cancer in a patient sample, and
CC for detecting the presence or absence of a polynucleotide encoded by a
```

CC nucleic acid which hybridises to (I) in a cell. A probe/primer derived  
CC from (I) can be used for determining the presence of a nucleic acid which  
CC hybridises to (I), and for determining the phenotype of cells in a sample  
CC of cells from a patient. (I) is useful for determining the presence of  
CC colon cancer in a cell or tissue type, for determining the presence of  
CC state of other type of cancer, in antisense therapy, to generate  
CC macroarrays on a solid surface, to identify a chromosome on which the  
CC corresponding gene resides, and in tissue profiling, forensics, genetic  
CC analysis, mapping and diagnostic applications. (I) can be used to raise  
CC antibodies, and to screen for peptide analogues and antagonists  
XX

SQ Sequence 620 BP; 152 A; 154 C; 161 G; 142 T; 0 U; 11 Other;

Query Match 100.0%; Score 42; DB 6; Length 620;  
Best Local Similarity 100.0%; Pred. No. 1.3e-07; Length 620;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACGCTTGGCCGAGCCCTTACAAGAGCTATGAAGTAAAG 42  
|||||  
Db 154 GCCACGCTTGGCCGAGCCCTTACAAGAGCTATGAAGTAAAG 113

## RESULT 7

ABQ59698/c  
ID ABQ59698 standard; cDNA; 634 BP.

XX ABQ59698;

DT 02-AUG-2002 (first entry)

XX Human colon cancer related nucleotide sequence SEQ ID NO:3393.

XX Human; colon cancer; cancer; tissue profiling; forensics; mapping;

KW Genetic analysis; diagnostic; antisense therapy; gene; ss.

XX Homo sapiens.

OS WO200229086-A2.

PN 11-APR-2002.

PF 02-OCT-2001; 2001WO-US030732.

PR 02-OCT-2000; 2000US-0237271P.

XX (FARB ) BAYER CORP.

XX Burgess C, Astle JH, Carroll E, Catino TJ, Dwivedi P, Molino GA;  
PI Thianglam A, Lewis ME;

XX WPI; 2002-426115/45.

XX New isolated nucleic acid that is differentially expressed in cancer  
PT tissues useful for determining the presence of colon cancer in a cell or  
PT tissue type, and in antisense therapy.

PS Claim 1; Fig 1; 796pp; English.

XX ABQ56306 to ABQ60787 represent isolated nucleic acids (I) differentially  
CC expressed in cancer tissues. ABQ78993 to ABQ79004 represent proteins  
CC encoded by the ABQ60776 to ABQ60787 nucleic acid sequences. (I) can be  
CC used in antisense therapy. An antibody immunoreactive with a polypeptide  
CC encoded by (I) is useful for detecting cancer in a patient sample, and  
CC for detecting the presence or absence of a polynucleotide encoded by a  
CC nucleic acid which hybridises to (I) in a cell. A probe/primer derived  
CC from (I) can be used for determining the presence of a nucleic acid which  
CC hybridises to (I), and for determining the phenotype of cells in a sample  
CC of cells from a patient. (I) is useful for determining the presence of  
CC colon cancer in a cell or tissue type, for determining the presence or  
CC state of other type of cancer, in antisense therapy, to generate  
CC macroarrays on a solid surface, to identify a chromosome on which the  
CC corresponding gene resides, and in-tissue profiling, forensics, genetic  
CC analysis, mapping and diagnostic applications. (I) can be used to raise

CC antibodies, and to screen for peptide analogues and antagonists  
XX

SQ Sequence 634 BP; 153 A; 154 C; 168 G; 146 T; 0 U; 13 Other;

Query Match 100.0%; Score 42; DB 6; Length 634;  
Best Local Similarity 100.0%; Pred. No. 1.3e-07;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACGCTTGGCCGAGCCCTTACAAGAGCTATGAAGTAAAG 42  
|||||  
Db 158 GCCACGCTTGGCCGAGCCCTTACAAGAGCTATGAAGTAAAG 117

## RESULT 8

ABT22328/c  
ID ABT22328 standard; DNA; 642 BP.

XX AC ABT22328;

DT 16-APR-2003 (first entry)

XX Breast cancer marker gene SEQ ID No 701.

XX Cytostatic; vaccine; breast cancer marker gene; breast mass; immunogen;  
KW chemotherapy; tumour burden; bait protein; two-hybrid; three-hybrid;  
KW surrogate marker gene; pharmacodynamic marker gene; transgenic animal;  
KW human; ds.

XX Homo sapiens.

OS WO200285298-A2.

PN 31-OCT-2002.

PF 19-APR-2002; 2002WO-US012612.

PR 20-APR-2001; 2001US-0285163P.

XX (MILL-) MILLENNIUM PHARM INC.

XX Lillie J, Palermo A, Wang Y, Steinmann K, Elias J, Mertens M;

XX WPI; 2003-093053/08.

XX Novel isolated polypeptide encoded by breast cancer marker gene, useful  
PT for diagnosing, staging, monitoring, prognosing and treating diseases  
PT associated with breast cancer.

PS Disclosure; Page 187-188; 725pp; English.

XX The invention relates to an isolated polypeptide encoded by a breast  
CC cancer marker gene comprising any of 141/21-805 nucleotide sequences,  
CC given in the specification. The methods of the invention are useful for  
CC diagnosing patients having an identified breast mass or symptoms  
CC associated with breast cancer, to diagnose breast cancer or its  
CC precursors, and for monitoring the efficacy of treatment of a breast  
CC cancer patient (e.g. efficacy of chemotherapy). The methods are also  
CC useful for evaluating a patient before, after or during therapy, to  
CC evaluate the reduction in a tumour burden. The breast cancer marker gene  
CC proteins are useful as immunogens for raising antibodies, by immunising a  
CC mammal with a breast cancer marker protein. The marker proteins are  
CC useful as bait proteins in a two-hybrid or three-hybrid assay, to  
CC identify other proteins which bind to or interact with the marker  
CC proteins. The breast cancer marker genes are useful as surrogate marker  
CC genes for one or more disorders, disease states or conditions leading to  
CC disease states, in particular, breast cancers. The breast cancer marker  
CC genes are useful as pharmacodynamic marker genes. An antibody which  
CC selectively binds to a protein of a breast cancer marker gene is useful  
CC for treating cancers, particularly breast cancers. The host cell of the  
CC polynucleotide sequence represents one of the breast cancer marker genes  
CC of the invention





CC dermal papilla, human keratinocytes and neonatal foreskin fibroblasts,  
CC and mouse embryonic skin, keratinocyte stem cells and transit amplifying  
CC cells. Polypeptides of the invention may be used to treat inflammation,  
CC cancer and neurological diseases. The proteins may be used to stimulate  
CC the growth and motility of keratinocytes, to inhibit the growth of cancer  
CC cells, to modulate angiogenesis and tumour vascularisation, to modulate  
CC skin inflammation, to modulate epithelial cell growth and to inhibit  
CC binding of HIV-1 to leukocytes. The invention may also be used to treat  
CC growth and developmental defects, skin wounds and hair follicle  
CC disorders. Sequences AAZ61606-261832 represent cDNA sequences derived  
CC from several mice, rat or human skin cell types. Sequences AAZ61606-  
CC Z61649, AAZ61725-261765, AAZ61802-261811 and AAZ61826 encode proteins  
CC with an N-terminal signal sequence, indicating that the proteins are  
CC secreted. Sequences AAZ61650-261668, AAZ61766-261780, AAZ61812-261817 and  
CC AAZ61827-261829 encode proteins with one or more putative transmembrane  
CC domains

XX  
SQ Sequence 1212 BP; 242 A; 350 C; 312 G; 308 T; 0 U; 0 Other;  
Query Match 100.0%; Score 42; DB 3; Length 1212;  
Best Local Similarity 100.0%; Pred. No. 1.5e-07; Mismatches 0; Indels 0; Gaps 0;  
Matches 42; Conservative 0;

QY 1 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
|||||  
Db 990 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1031

## RESULT 11

AAC99709

ID AAC99709 standard; cDNA; 1212 BP.

XX AAC99709;

AC AAC99709;

XX 08-MAR-2001 (first entry)

DT 08-MAR-2001 (first entry)

XX Skin cell cDNA, SEQ ID NO: 249.

DE

XX Human; skin cell; cytostatic; antiinflammatory; anti-HIV; nootropic;  
KW neuroprotective; vulnery; immunomodulatory; vaccine;  
KW keratinocyte growth stimulation; cancer; angiogenesis inhibition;  
KW inflammation; neurological disease; ss.

XX Homo sapiens.

OS

XX Watson JD, Strachan L, Sleeman M, Kumble KD;  
FN Murison JG;  
XX WO200069884-A2.

XX 23-NOV-2000.

PD

XX 15-MAY-2000; 2000WO-NZ000075.

PF

XX 14-MAY-1999; 99US-00312283.

PR

XX (GENE-) GENESIS RES &amp; DEV CORP LTD.

PA

XX Watson JD, Strachan L, Orrust R, Murison JG;  
PI Murison JG;  
XX Kumble KD;

XX WPI; 2001-007495/01.

XX

DR P-PSDB; AAB55958.

DR

XX New isolated polynucleotide used in the identification of genetic  
PT disorders and encoding polypeptides used for treating inflammatory  
PT disease, cancer and neurological diseases.

PT

XX Claim 1; Page 210-211; 352pp; English.

PS

XX The present polynucleotide encodes a polypeptide which is expressed in  
CC mammalian skin cells. The polypeptide is useful for stimulating  
CC keratinocyte growth and motility, inhibiting the growth of cancer cells,  
CC modulating angiogenesis, inhibiting angiogenesis and vascularisation of  
CC tumours, modulating skin inflammation, stimulating the growth of  
CC epithelial cells, inhibiting the binding of human immunodeficiency virus

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC (HIV)-1 to leukocytes, and treating inflammatory disease, cancer and  
CC neurological diseases. The polynucleotide can be used as a marker, in the  
CC identification of genetic disorders, and for the design of  
CC oligonucleotides for examining expression patterns

XX Sequence 1212 BP; 242 A; 350 C; 312 G; 308 T; 0 U; 0 Other;  
Query Match 100.0%; Score 42; DB 4; Length 1212;  
Best Local Similarity 100.0%; Pred. No. 1.5e-07; Mismatches 0; Indels 0; Gaps 0;  
Matches 42; Conservative 0;

QY 1 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42

|||||

Db 990 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1031

## RESULT 12

ABL34861

ID ABL34861 standard; cDNA; 1212 BP.

XX ABL34861;

AC ABL34861;

XX 04-APR-2002 (first entry)

DT 04-APR-2002 (first entry)

XX Human cDNA isolated from skin cells SEQ ID NO: 249.

DE

XX Human; rat; mouse; skin cell; skin wound; cancer; growth defect;  
KW developmental defect; inflammatory disease; dermatological; vulnery;  
KW immunomodulator; anti-inflammatory; cytostatic; neuroprotective; gene;  
KW ss.

XX Homo sapiens.

OS

XX WO200190357-A1.

PN

XX 29-NOV-2001.

PD

XX 24-MAY-2001; 2001WO-NZ000099.

PF

XX 24-MAY-2000; 2000US-0206650P.

PR

XX 25-JUL-2000; 2000US-0221232P.

PR

XX (GENE-) GENESIS RES &amp; DEV CORP LTD.

PA

XX Watson JD, Strachan L, Sleeman M, Orrust R, Murison JG;  
PI Kumble KD;

PI

XX WPI; 2002-122020/16.

DR

XX New polynucleotides and polypeptides encoded by the polynucleotides  
PT isolated from skin cells, useful for treating skin wounds, cancers,  
PT growth and developmental defects, inflammatory diseases, or for  
PT modulating immune responses.

PT

XX Claim 1; Page 175; 466pp; English.

PS

XX The present invention provides the protein and coding sequences of cDNAs  
CC isolated from human, murine and rat skin cell libraries. The sequences  
CC can be used in the development of therapeutic agents useful in the  
CC treatment of skin diseases, including skin wounds, cancer, growth  
CC defects, developmental defects and inflammatory diseases. The proteins  
CC have important roles in the induction of hair growth, cell proliferation  
CC and cell-cell interaction, in maintaining tissue integrity, in wound  
CC healing and in modulating immune responses. The present sequence is a  
CC cDNA of the invention

CC

XX Sequence 1212 BP; 242 A; 350 C; 312 G; 308 T; 0 U; 0 Other;

SQ

Query Match 100.0%; Score 42; DB 6; Length 1212;

Best Local Similarity 100.0%; Pred. No. 1.5e-07; Mismatches 0; Indels 0; Gaps 0;  
Matches 42; Conservative 0;

QY 1 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42

```
|||||
990 GCCCAGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 1031

RESULT 13
AC04775
ID ACA04775 standard; cDNA; 1228 BP.
AC ACA04775;
XX
XX
XX
XX 28-MAY-2003 (first entry)
XX
XX cDNA encoding human membrane associated protein fragment #223.
XX
XX Human; ss; gene; microarray; membrane-associated protein; neuropathology;
KW immunopathology; pancreatic disease; cancer; diabetes; hyperlipidaemia;
KW pancreatic cholera; Alzheimer's disease; Huntington's disease; sarcoma;
KW fibrocystic disease; leukaemia; adenocarcinoma; AIDS; allergy; anaemia;
KW asthma; gout; dementia.
XX
XX Homo sapiens.
OS
XX US6492505-B1.
PN
XX
XX 10-DEC-2002.
PD
XX
XX 31-JAN-2000; 2000US-00495050.
PF
XX
XX 01-FEB-1999; 99US-0118318P.
PR
XX (INCY-) INCYTE GENOMICS INC.
PA
XX Reddy R, Guegler KJ, Au-Young J;
PI
XX WPI; 2003-327324/31.
DR
XX
XX Combination for research/diagnostic applications and for monitoring
PT treatment of e.g., cancer, comprises polynucleotides comprising a
PT fragment of gene encoding membrane-associated proteins, receptors or ion
PT channels.
PS
XX Claim 1; Col 215-216; 147pp; English.
XX
XX The invention relates to a combination comprising several polynucleotide
CC sequences comprising a fragment of gene encoding membrane-associated
CC proteins, receptors or ion channels. The combination is useful as a
CC probe, for research and diagnostic applications, for monitoring the
CC expression of several expressed polynucleotides, in the diagnosis and
CC monitoring of treatment of pancreatic disease, cancer, immunopathology or
CC neuropathology, for investigating an individual's predisposition to the
CC above disease, in genetic or gene expression analysis of polynucleotide
CC sequences, to investigate cellular responses to infection or drug
CC treatment, as hybridisable array elements in a microarray, to purify a
CC subpopulation of mRNAs, cDNAs or genomic fragments in a sample, in
CC diagnostics, prognostics and treatment regimens, in drug discovery and
CC development, in toxicological and carcinogenicity studies, and in
CC forensics or pharmacogenomics, to monitor the progression of disease, to
CC monitor the efficacy of treatment, to diagnose the conditions of the
CC pancreas e.g. diabetes, pancreatic cholera, hyperlipidaemia or
CC fibrocystic disease, to diagnose a cancer e.g. leukaemia, adenocarcinoma
CC or sarcoma, to diagnose immunopathologies e.g. AIDS, allergies, anaemia,
CC asthma or gout, to diagnose neuropathologies e.g. Alzheimer's disease,
CC dementia or Huntington's disease, to rapidly screen large numbers of
CC candidate drug molecules and as query sequences against GenBank,
CC SwissProt, BLOCKS and PRINTS databases. The combination is employed to
CC fine tune the treatment regimen and thus the expression patterns
CC associated with undesirable side effects are avoided. The present
CC sequence represents a cDNA encoding a fragment of gene encoding human
CC membrane-associated proteins, receptors or ion channels
XX
XX Sequence 1228 BP; 272 A; 341 C; 283 G; 332 T; 0 U; 0 Other;
SQ
XX Query Match 100.0%; Score 42; DB 8; Length 1228;
```

```
Best Local Similarity 100.0%; Pred. No. 1.5e-07;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCCAGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 42
   |||||||
Db 541 GCCCAGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 582

RESULT 14
AEA00112
ID AEA00112 standard; cDNA; 1460 BP.
XX
XX AEA00112;
XX
XX 28-JUL-2005 (first entry)
XX
XX Human TAT143 cDNA SEQ ID NO:64.
DE
XX
XX ss; gene; tumor-associated antigen; cytostatic; breast tumor;
KW endocrine disease; gynecology and obstetrics; neoplasm; colon tumor;
KW gastrointestinal disease; rectal tumor; endometroid carcinoma;
KW genitourinary disease; renal tumor; lung tumor; respiratory disease;
KW ovary tumor; skin tumor; liver tumor.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH 128..1201
CDS /*tag= a
   /*product= "TAT143"
FT
FT
XX
XX US2005106644-A1.
PN
XX
XX 19-MAY-2005.
PD
XX
XX 08-SEP-2004; 2004US-00936626.
PF
XX
XX 20-JUN-2001; 2001US-0299500P.
PR
XX 29-JUN-2001; 2001US-0301880P.
PR
XX 18-SEP-2001; 2001US-0323268P.
PR
XX 19-JUN-2002; 2002US-00177488.
PR
XX 26-MAR-2004; 2004US-0557116P.
PR
XX 04-AUG-2004; 2004US-0598899P.
XX
XX (GETH ) GENENTECH INC.
PA
XX
XX Cairns B, Chen R, Frantz G, Hillan KJ, Koeppen H, Phillips HS;
PI Polakis P, Spencer SD, Smith V, Williams PM, Wu TD, Zhang Z;
PI Sliwkowski M;
PI
XX
XX WPI; 2005-384304/39.
DR
XX P-PSDB; AEA00190.
DR
XX
XX Novel isolated antibody capable of binding to tumor-associated antigenic
PT target polypeptide, useful for treating cell proliferative disorder e.g.
PT cancer.
XX
XX Claim 1; SEQ ID NO 64; 337pp; English.
PS
XX
XX The invention relates to a novel isolated antibody binding to a
CC polypeptide having at least 80% sequence identity to a polypeptide having
CC any one of 76 fully defined 182-910 amino acid tumor-associated antigenic
CC target polypeptide (TAT) sequences (AEA00127-AEA00202) given in the
CC specification, a polypeptide having any one of (AEA00127-AEA00202),
CC lacking its associated signal peptide, or an extracellular domain of a
CC polypeptide having any one of (AEA00127-AEA00202). The polypeptide is
CC encoded by the nucleotide sequence having any one of (AEA00049-AEA00126).
CC An antibody of the invention has cytostatic activity. The antibody is
CC useful for inhibiting growth of a cell expressing TAT188, which involves
CC contacting the cell with the antibody. The cell is a cancer cell chosen
CC from breast, colon, rectum, endometrium, kidney, lung, ovary, skin and
CC liver cell. The cancer cell is a mammalian cell, preferably a human cell.
CC The antibody is also useful for detecting the level of TAT188
```

CC polypeptide expressed in a test cell relative to a control cell, and for  
CC detecting the level of TAT188 polypeptide or a polypeptide having at  
CC least 80 % sequence identity to the TAT188 polypeptide sequence in a test  
CC cell relative to a control cell. The antibody is useful for treating a  
CC preventing a cell proliferative disorder associated with increased  
CC expression or activity of a polypeptide having at least 80 % identity to  
CC a TAT188 polypeptide sequence. The cell proliferative disorder is cancer.  
CC The method of the invention is useful for inhibiting the growth of a  
CC cancer cell. The present sequence encodes a polypeptide of the invention.  
XX  
SQ Sequence 1460 BP; 302 A; 418 C; 362 G; 378 T; 0 U; 0 Other;

Query Match 100.0%; Score 42; DB 14; Length 1460;  
Best Local Similarity 100.0%; Pred. No. 1.5e-07;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
|||||  
Db 1145 GCCCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1186

## RESULT 15

AEA00632  
ID AEA00632 standard; cDNA; 1460 BP.

XX  
AC AEA00632;

DT 28-JUL-2005 (first entry)

DE Human TAT143 cDNA sequence SeqID64.

XX antibody identification; tumor-associated antigen; cytostatic;  
KW RNA interference; gene therapy; cell death; cancer; breast tumor;  
KW colon tumor; rectal tumor; renal tumor; lung tumor; ovary tumor;  
KW skin tumor; liver tumor; gene; ss; TAT143.

XX Homo sapiens.

PN US2005107595-A1.

XX 19-MAY-2005.

PF 10-SEP-2004; 2004US-00938061.

XX 20-JUN-2001; 2001US-0299500P.

PR 29-JUN-2001; 2001US-0301880P.

PR 18-SEP-2001; 2001US-0323268P.

PR 19-JUN-2002; 2002US-00177488.

PR 26-MAR-2004; 2004US-0557116P.

PR 04-AUG-2004; 2004US-0598899P.

PA (GETH ) GENENTECH INC.

XX Cairns B, Chen R, Frantz G, Hillan KJ, Koeppen H, Phillips HS;

PI Polakis P, Spencer SD, Smith V, Williams PM, Wu TD, Zhang Z;

PI Sakanaka C, Chuntharapai A, Reed CJ;

XX WPI; 2005-371577/38.

DR P-PSDB; AEA00710.

XX Novel isolated antibody e.g. anti-E16 or anti-TAT112 antibody that binds  
PT to tumor-associated antigenic target polypeptide, useful for diagnosing  
PT or treating cancer.

XX Claim 1; SEQ ID NO 64; 96pp; English.  
XX This invention relates to a novel isolated antibody, for example anti-E16  
CC or anti-TAT112 antibody, that binds to a tumor-associated antigenic  
CC target polypeptide (TAT) and that lacks an associated signal peptide  
CC sequence. The invention may be useful for the development of compounds  
CC with a cytostatic activity acting as antagonists of the TAT188  
CC polypeptide or RNA interference whilst the disclosed sequences may be  
CC useful for gene therapy. The invention is useful for inducing the death

CC of a cell (such as a cancer cell chosen from breast, colon, rectum,  
CC endometrium, kidney, lung, ovary, skin and liver) to which it binds,  
CC inhibiting proliferation or promoting cell death of a cell expressing  
CC TAT188. In addition, the invention may be useful for detecting the level  
CC of TAT188 polypeptide in a test cell relative to a control cell, or  
CC treating or preventing a cell proliferative disorder associated with  
CC increased expression of TAT188. The novel antibody of the invention is  
CC useful for inhibiting the growth of a cancer cell and may be useful for  
CC diagnosing or treating cancer. The present sequence is that of the human  
CC TAT143 cDNA which encodes a protein against which an antibody of the  
CC invention may be targeted.  
XX

SQ Sequence 1460 BP; 302 A; 418 C; 362 G; 378 T; 0 U; 0 Other;

Query Match 100.0%; Score 42; DB 14; Length 1460;  
Best Local Similarity 100.0%; Pred. No. 1.5e-07;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
|||||  
Db 1145 GCCCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1186

## RESULT 16

AAF58615

ID AAF58615 standard; cDNA; 1619 BP.

XX AAF58615;

DT 24-APR-2001 (first entry)

DE Human RECAP polynucleotide, SEQ ID NO: 43.

XX Human; RECAP; receptors and associated proteins; cerebroprotective;  
KW nootropic; neuroprotective; anticonvulsant; antiparkinsonian; anti-HIV;  
KW antidiabetic; immunostimulant; immunomodulator; antiinflammatory;  
KW antithyroid; immunosuppressive; nephrotropic; antigout; thyromimetic;  
KW cytostatic; antibacterial; virucide; fungicide; protozoacide;  
KW antiarteriosclerotic; hepatotropic; gene therapy; infection; cancer; ss.

XX Homo sapiens.

PN WO200107612-A2.

PD 01-FEB-2001.

PF 21-JUL-2000; 2000WO-US020035.

XX 21-JUL-1999; 99US-0145232P.

PR 07-OCT-1999; 99US-0158578P.

PR 12-NOV-1999; 99US-0165192P.

XX (INCY-) INCYTE GENOMICS INC.

XX Au-Young J, Bandman O, Tang YT, Yue H, Azimzai Y, Burford N;

PI Baughn MR, Lu DAM, Hillman JL, Patterson C, Lal P;

XX WPI; 2001-168554/17.

DR P-PSDB; AAB68891.

XX Novel receptors and associated proteins for diagnosis and treatment of  
PT neurological disorders, immunological disorders including autoimmune/  
PT inflammatory disorders and cell proliferative disorders such as cancer.

XX Example 5; Page 126-127; 128pp; English.

XX The present sequence encodes a human RECAP (receptors and associated  
CC proteins) polypeptide. RECAP polynucleotides and polypeptides are useful  
CC in the diagnosis, treatment and prevention of neurological disorders such  
CC as stroke, Alzheimer's disease, Pick's disease, Huntington's disease,  
CC dementia, Parkinson's disease, Down's syndrome, amyotrophic lateral  
CC sclerosis, multiple sclerosis, bacterial and viral meningitis, CJD  
CC (Creutzfeldt-Jakob disease), GSS (Gerstmann -Straussler-Scheinker

```
CC syndrome); immunological disorders, including autoimmune/inflammatory
CC disorders such as AIDS, DiGeorge's syndrome, severe combined
CC immunodeficiency disease (SCID), Chediak-Higashi syndrome, Cushing's
CC disease, Addison's disease, autoimmune thyroiditis, Crohn's disease,
CC diabetes mellitus, Good pasture's syndrome, gout, Grave's disease,
CC Hashimoto's thyroiditis, Sjogren's syndrome, Werner's syndrome, viral,
CC bacterial, fungal, parasitic, protozoal, and helminthic infections; and
CC cell proliferation disorders such as arteriosclerosis, atherosclerosis,
CC cirrhosis, hepatitis and cancer
XX
SQ Sequence 1619 BP; 331 A; 463 C; 400 G; 425 T; 0 U; 0 Other;
Query Match 100.0%; Score 42; DB 4; Length 1619;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCCCAGCGTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42
DB 1137 GCCCAGCGTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1178
RESULT 17
AAZ90056
ID AAZ90056 standard; cDNA; 1718 BP.
XX
AC AAZ90056;
XX
DT 09-MAY-2000 (first entry)
XX
DE Hydrophobic domain containing protein clone HP10549 nucleotide sequence.
XX
KW Hydrophobic domain; clone HP10549; nutritional supplement; SCID; HIV;
KW cell proliferation; immune stimulant; immune deficiency; tumour; pain;
KW rheumatoid arthritis; insulin dependent diabetes mellitus; fertility;
KW myasthenia gravis; haematopoiesis regulator; tissue growth; depression;
KW anti-inflammatory; infection; bodily characteristic; ss.
XX
OS Homo sapiens.
XX
PN WO200000506-A2.
XX
PD 06-JAN-2000.
XX
PF 18-JUN-1999; 99WO-JP003242.
XX
PR 26-JUN-1998; 98JP-00180008.
XX
PA (SAGA) SAGAMI CHEM RES CENT.
PA (PROT-) PROTEGENE INC.
XX
PI Kato S, Kimura T;
XX
DR WPI; 2000-160665/14.
DR P-PSDB; AAY78809.
XX
PT Novel human proteins having hydrophobic domains used for research and
PT diagnostic purposes.
XX
PS Claim 4; Page 111-113; 117pp; English.
XX
CC This sequence represents the hydrophobic domain containing protein, clone
CC HP10549 nucleotide sequence. The sequence is isolated from a human
CC stomach cancer cell line. The invention relates to human proteins with
CC hydrophobic domains, the DNA and the cDNA encoding them. The
CC polynucleotides and proteins are predicted to have biological activities
CC which make them suitable for treating, preventing or ameliorating medical
CC conditions in humans and animals. Suggested activities include
CC nutritional activity (nutritional source or supplement); cytokine and
CC cell proliferation/differentiation activity; immune stimulating (e.g. as
CC vaccines) or suppressing activity (e.g. to treat various immune
CC deficiencies such as SCIPs or HIV, connective tissue disease, systemic
CC lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary
CC inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin
CC
```

```
CC dependent diabetes mellitus, myasthenia gravis, graft-versus-host disease
CC and autoimmune inflammatory eye disease, as well as asthma, allergies and
CC organ transplantation; haematopoiesis regulating activity (e.g. in
CC treatment of myeloid or lymphoid cell deficiencies); tissue growth
CC activity (e.g. wound healing and tissue repair, ulcers, burns,
CC periodontal disease); activin/inhibin activity; chemotactic/chemokinetic
CC activity; haemostatic and thrombolytic activity (e.g. treating
CC haemophilias); receptor/ligand activity; anti-inflammatory activity; and
CC tumour inhibition activity. The polynucleotides are also stated to be
CC useful for gene therapy. Other activities include inhibiting infections
CC caused by bacteria, fungi, viruses and other parasites (e.g. Hepatitis,
CC malaria); effecting bodily characteristics such as, e.g. weight, colour,
CC skin, effecting biorhythms or cardiac cycles; enhancing fertility;
CC treatment of depression; treatment of pain; hormonal or endocrine
CC activity. The polynucleotides may also be used for recombinant expression
CC of the protein
XX
SQ Sequence 1718 BP; 381 A; 495 C; 377 G; 465 T; 0 U; 0 Other;
Query Match 100.0%; Score 42; DB 3; Length 1718;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCCCAGCGTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42
DB 561 GCCCAGCGTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 602
RESULT 18
ABT10173
ID ABT10173 standard; cDNA; 2302 BP.
XX
AC ABT10173;
XX
DT 04-DEC-2002 (first entry)
XX
DE Human breast cancer associated coding sequence SEQ ID NO: 307.
XX
OS Homo sapiens.
XX
PN WO200259271-A2.
XX
PD 01-AUG-2002.
XX
PF 25-JAN-2002; 2002WO-US002176.
XX
PR 25-JAN-2001; 2001US-0263757P.
PR 25-APR-2001; 2001US-0286090P.
PR 23-MAY-2001; 2001US-0292517P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Orr MS, Nation M, Diggans JC, Zeng W;
XX
DR WPI; 2002-674803/72.
XX
PT Diagnosing breast cancer in a patient comprises detecting the level of
PT gene expression in cell or tissue samples, where a differential gene
PT expression is indicative of breast cancer.
XX
PS Claim 1; SEQ ID NO 307; 260pp + Sequence Listing; English.
XX
CC The present invention relates to methods of diagnosing breast cancer in a
CC patient, which comprise detecting the level of expression in a tissue
CC sample of two or more genes selected from those shown in ABT09867-
CC ABT1112, where a differential expression of the genes indicates breast
CC cancer. The methods are useful in diagnosing, treating, detecting the
CC progression, and in monitoring treatment of breast cancer in patients.
CC The methods are also useful as a screening tool for agents that modulate
CC the onset or progression of breast cancer. The breast cancer genes may be
CC
```

CC used as diagnostic markers for the prediction or identification of the  
CC malignant state of breast tissue, for confirming the type and progression  
CC of cancer, and for drug screening and assays. The present sequence is a  
CC coding sequence of the invention. Note: The sequence data for this patent  
CC did not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub.published\_pct\_sequences

SQ Sequence 2302 BP; 494 A; 666 C; 533 G; 609 T; 0 U; 0 Other;

Query Match 100.0%; Score 42; DB 6; Length 2302;  
Best Local Similarity 100.0%; Pred. No. 1.7e-07;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACGCTTGCGCGAGCCCTTACAAGACTATGAAGTAAAG 42  
|||||  
DB 1117 GCCACGCTTGCGCGAGCCCTTACAAGACTATGAAGTAAAG 1158

## RESULT 19

ACC58386  
ID ACC58386 standard; cDNA; 2302 BP.

XX AC ACC58386;

XX 26-AUG-2003 (first entry)

XX Human GPCR-like retinoic acid-induced gene 1.

XX Human; retinoic acid-induced gene 1; RAIG1; feeding; fasting; GPCR;  
XX receptor; G-protein coupled receptor; anorectic; antidiabetic;  
XX antidepressant; immunomodulator; transgenic; gene therapy; gene; ss.  
XX Homo sapiens.

XX Key Location/Qualifiers  
FH Key 100.1173  
FT CDS /\*tag= a  
FT /\*product= "GPCR-like RAIG1"

XX WO2003016553-A2.

XX 27-FEB-2003.

XX 20-AUG-2002; 2002WO-US026510.

XX 20-AUG-2001; 2001US-0313940P.

XX (GETH ) GENENTECH INC.  
XX (CURA-) CURAGEN CORP.

XX Lewin DA, Stewart TA;

XX WPI; 2003-278580/27.

XX P-PSDB; ABR42649.

XX New G-protein coupled receptor-like retinoic acid induced gene 1 (GPCR-  
XX like RAIG1) polypeptide and gene, useful for diagnosing or treating  
XX metabolic disorders, e.g. obesity, anorexia, cachexia or diabetes.

XX Claim 10; Page 16-17; 150pp; English.

XX The present sequence is that of human G-protein coupled receptor-like  
XX retinoic acid induced gene 1 (GPCR-like RAIG1). This is the human  
XX homologue of murine GPCR-like RAIG1. The murine gene was shown to be  
XX differentially regulated during fasting-feeding cycles in mice, with  
XX moderate induction early in fasting, down-regulation with extended  
XX fasting and 4-fold up-regulation with feeding in recovery from fasting.  
XX The differentially expressed gene, its mRNA, and the encoded protein, can  
XX each be manipulated to detect and treat metabolic disorders associated  
XX with up- or down-regulation of GPCR-like RAIG1 activity, such as obesity,  
XX anorexia, cachexia or diabetes

SQ Sequence 2302 BP; 494 A; 666 C; 533 G; 609 T; 0 U; 0 Other;

Query Match 100.0%; Score 42; DB 8; Length 2302;  
Best Local Similarity 100.0%; Pred. No. 1.7e-07;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACGCTTGCGCGAGCCCTTACAAGACTATGAAGTAAAG 42  
|||||  
DB 1117 GCCACGCTTGCGCGAGCCCTTACAAGACTATGAAGTAAAG 1158

## RESULT 20

ADD93240  
ID ADD93240 standard; cDNA; 2302 BP.

XX AC ADD93240;

XX 29-JAN-2004 (first entry)

XX RAIG1 coding sequence.

XX ss; gene; retinoic acid-inducible gene 1; RAIG1; orphan;  
XX G-protein coupled receptor; GPCR; chromosome 12; GPCR5B; carcinoma;  
XX vaccine; breast cancer; pancreatic cancer; lung cancer; liver cancer;  
XX ovarian cancer; colon cancer; osteosarcoma.

XX Homo sapiens.

XX Key Location/Qualifiers  
FH Key 100.1173  
FT CDS /\*tag= a  
FT /\*product= "RAIG1"

XX WO2003087832-A2.

XX 23-OCT-2003.

XX 10-APR-2003; 2003WO-GB001587.

XX 11-APR-2002; 2002GB-00008331.

XX 17-SEP-2002; 2002GB-00021538.

XX (OXFO-) OXFORD GLYSCSCIENCES UK LTD.

XX Terrett JA;

XX WPI; 2003-845382/78.

XX P-PSDB; ADD93239.

XX GENBANK; AF095448.

XX Screening, diagnosing and/or treating carcinoma, including breast,  
XX pancreatic, lung, liver, ovarian and colon cancer by detecting the change  
XX in expression or activity of an RAIG1 polypeptide or encoding nucleic  
XX acid molecule.

XX Claim 1; Fig 2; 43pp; English.

XX This sequence encodes retinoic acid-inducible gene 1 (RAIG1) polypeptide.  
XX RAIG1 is an orphan G-protein coupled receptor (GPCR) located on  
XX chromosome 12. RAIG1 shows a restricted expression pattern compared to a  
XX related receptor, GPCR5B, which is widely expressed in peripheral and  
XX central tissues. The RAIG1 cDNA and polypeptide sequence may be used in  
XX the method of the invention for screening for and/or diagnosis of  
XX carcinoma in a subject, and/or monitoring the effectiveness of carcinoma  
XX therapy. The method comprises detecting and/or quantifying in a  
XX biological sample obtained from the subject an RAIG1 polypeptide and a  
XX nucleic acid molecule. The RAIG1 polypeptide and nucleic acid molecule  
XX are useful in the manufacture of a medicament for the treatment of  
XX carcinoma, where the composition is a vaccine. An agent which interacts  
XX with or causes change in the expression or activity of an RAIG1  
XX polypeptide or nucleic acid molecule, is also useful in the manufacture  
XX of a medicament for the treatment of carcinoma that is breast cancer,  
XX pancreatic cancer, lung cancer, liver cancer, ovarian cancer, colon

CC cancer and/or osteosarcoma. They can also be used in the diagnosis and  
CC screening of such carcinomas.  
SQ Sequence 2302 BP; 494 A; 666 C; 533 G; 609 T; 0 U; 0 Other;  
Query Match 100.0%; Score 42; DB 10; Length 2302;  
Best Local Similarity 100.0%; Pred. No. 1.7e-07; Mismatches 0; Indels 0; Gaps 0;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GCCCAGCGTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
Db 1117 GCCCAGCGTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1158  
RESULT 21  
ADL24773  
ID ADL24773 standard; DNA; 2302 BP.  
XX  
AC ADL24773;  
XX  
DT 20-MAY-2004 (first entry)  
XX  
DE Intestinal epithelium/peyer's patch M cell-associated DNA sequence #171.  
XX  
KW intestinal epithelium cell development; peyer's patch M cell development;  
KW inflammatory bowel disease; glutenenteropathy; infectious disease;  
KW autoimmune disease; haemolytic anaemia; rheumatoid arthritis; dermatitis;  
KW Grave's disease; multiple sclerosis; allergy; asthma; diabetic mellitus;  
KW immune system disorder; hypersensitivity; anaphylaxis;  
KW blood group incompatibility; ds; human.  
XX  
OS Homo sapiens.  
XX  
PN WO200280852-A2.  
XX  
PD 17-OCT-2002.  
XX  
PF 04-APR-2002; 2002WO-US010873.  
XX  
PR 04-APR-2001; 2001US-0281416P.  
XX  
PA (DIGI-) DIGITAL GENE TECHNOLOGIES INC.  
XX  
PI Brayden DJ, Byrne D, O'mahony DJ, Evans CF, Mah SP, Lo DD;  
XX  
DR WPI; 2003-075470/07.  
XX  
PT Novel isolated or purified polypeptide encoded by genes associated with  
PT intestinal epithelium or M cell development, differentiation or function,  
PT useful for treating autoimmune diseases and infectious diseases.  
XX  
PS Claim 1; SEQ ID NO 283; 152pp; English.  
XX  
CC The invention comprises DNA sequences which are associated with  
CC intestinal epithelium and peyer's patch M cells. The DNA sequences of the  
CC invention are useful for assessing, modifying, modulating or regulating  
CC intestinal epithelium or M cell development. The DNA sequences of the  
CC invention are also useful in the treatment of: inflammatory bowel  
CC disease, glutenenteropathy, infectious diseases, autoimmune diseases  
CC (e.g. haemolytic anaemia, rheumatoid arthritis, dermatitis, Grave's  
CC disease, multiple sclerosis, allergy, asthma and diabetic mellitus),  
CC diseases or disorders of the immune system, hypersensitivity,  
CC anaphylaxis, and blood group incompatibility. The present nucleic acid  
CC represents an intestinal epithelium/peyer's patch M cell-associated DNA  
CC sequence of the invention. NOTE: The present sequence is not shown in the  
CC specification, but has been retrieved from the WIPO website.  
XX  
SQ Sequence 2302 BP; 494 A; 666 C; 533 G; 609 T; 0 U; 0 Other;  
Query Match 100.0%; Score 42; DB 10; Length 2302;  
Best Local Similarity 100.0%; Pred. No. 1.7e-07;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCCAGCGTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
Db 1117 GCCCAGCGTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1158  
RESULT 22  
ABT31923  
ID ABT31923 standard; DNA; 2316 BP.  
XX  
AC ABT31923;  
XX  
DT 01-MAY-2003 (first entry)  
XX  
DE Human breast cancer / ovarian cancer related coding sequence #30.  
XX  
KW Human; gene; ds; cytostatic; breast cancer; ovarian cancer.  
XX  
OS Homo sapiens.  
XX  
PN WO2003000012-A2.  
XX  
PD 03-JAN-2003.  
XX  
PF 21-JUN-2002; 2002WO-US019773.  
XX  
PR 21-JUN-2001; 2001US-0300159P.  
PR 27-JUN-2001; 2001US-0301351P.  
XX  
PA (MILL-) MILLENNIUM PHARM INC.  
XX  
PI Veiby OP;  
XX  
DR WPI; 2003-267848/26.  
DR P-PSDB; ABU37054.  
XX  
PT Determining the presence of breast cancer in an individual, involves  
PT using specific polynucleotide markers.  
XX  
PS Disclosure; Page 163; 233pp; English.  
XX  
CC The invention comprises a method for assessing whether a patient is  
CC afflicted with breast cancer or ovarian cancer. The method involves the  
CC use of specific DNA markers. The method of the invention is useful in the  
CC detection and treatment of ovarian and breast cancer. DNA sequences  
CC ABT31894 - ABT31949 encode human breast/ovarian cancer-related proteins  
XX  
SQ Sequence 2316 BP; 502 A; 670 C; 535 G; 609 T; 0 U; 0 Other;  
Query Match 100.0%; Score 42; DB 10; Length 2316;  
Best Local Similarity 100.0%; Pred. No. 1.7e-07; Mismatches 0; Indels 0; Gaps 0;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GCCCAGCGTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
Db 1123 GCCCAGCGTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1164  
RESULT 23  
AAH14688  
ID AAH14688 standard; cDNA; 2446 BP.  
XX  
AC AAH14688;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
DE Human cDNA sequence SEQ ID NO:12388.  
XX  
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX  
OS Homo sapiens.  
XX  
PN EP1074617-A2.  
XX

```

PD 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-00116126.
XX
PR 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX
XX (HELI-) HELIX RES INST.
PA
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT length cDNAs defined in the specification, and for the detection and/or
PT diagnosis of the abnormality of the proteins encoded by the full-length
PT cDNAs.
XX
XX Claim 8; SEQ ID NO 12388; 2537pp + Sequence Listing; English.
XX
XX The present invention describes primer sets for synthesizing 5602 full-
CC length cDNAs defined in the specification. Where a primer set comprises:
CC (a) an oligo-dT primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention
XX
SQ Sequence 2446 BP; 507 A; 723 C; 572 G; 644 T; 0 U; 0 Other;
Query Match 100.0%; Score 42; DB 4; Length 2446;
Best Local Similarity 100.0%; Pred. No. 1.7e-07;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCCACGCTTGGCGAGCCCTTACAAGACTATGAAGTAAAG 42
Db 1271 GCCACGCTTGGCGAGCCCTTACAAGACTATGAAGTAAAG 1312
RESULT 24
ABZ42832
ID ABZ42832 standard; DNA; 2456 BP.
XX
XX ABZ42832;
AC
XX 04-MAR-2003 (first entry)
DT
XX Human G protein-coupled receptor RAIG1 nucleotide SEQ ID NO:453.
DE
XX
XX G protein-coupled receptor; GPCR; antigenic peptide; gene therapy;
KW G protein-coupled receptor modulator; antibody; immune-related disease;
KW growth-related disease; cell regeneration-related disease; AIDS; cancer;
KW immunological-related cell proliferative disease; autoimmune disease;
KW Alzheimer's disease; atherosclerosis; infection; osteoarthritis; allergy;

```

---

```

KW osteoporosis; cardiomyopathy; inflammation; Crohn's disease; diabetes;
KW graft versus host disease; Parkinson's disease; multiple sclerosis; pain;
KW psoriasis; anxiety; depression; schizophrenia; dementia; memory loss;
KW mental retardation; epilepsy; asthma; tuberculosis; obesity; nausea;
KW hypertension; hypotension; renal disorder; rheumatoid arthritis; trauma;
KW ulcer; gene; ds.
XX Homo sapiens.
XX WO200261087-A2.
XX 08-AUG-2002.
XX
XX 19-DEC-2001; 2001WO-US050107.
XX
XX 19-DEC-2000; 2000US-0257144P.
XX
XX (LIFE-) LIFESPAN BIOSCIENCES INC.
PA
XX Burmer GC, Roush CL, Brown JP;
PI
XX
XX WPI; 2003-046718/04.
DR P-PSDB; ABP81984.
XX
XX New isolated antigenic peptides e.g., for G protein-coupled receptors
PT (GPCR), useful for diagnosing and designing drugs for treating conditions
PT in which GPCRs are involved, e.g. AIDS, Alzheimer's disease, cancer or
PT autoimmune diseases.
XX
XX Disclosure; Fig 1; 523pp; English.
XX
XX The present invention describes antigenic peptides (I) comprising: (a)
CC any one of 1601 sequences (see ABP82019 to ABP83619) of 12-24 amino
CC acids. Also described: (1) an assay for the detection of a particular G
CC protein-coupled receptor (GPCR) or a candidate polypeptide in a sample;
CC and (2) an isolated antibody having high specificity and high affinity or
CC avidity for a particular GPCR. (I) can be used as GPCR modulators and in
CC gene therapy. The antigenic peptides for GPCRs are useful in detecting an
CC antibody against a particular GPCR, and in the production of specific
CC antibodies. The peptides and antibodies are also useful for detecting the
CC presence or absence of corresponding GPCRs. The antigenic peptides for
CC GPCRs and antibodies are useful for diagnosing and designing drugs for
CC treating immune-related diseases, growth-related diseases, cell
CC regeneration-related disease, immunological-related cell proliferative
CC diseases, or autoimmune diseases, e.g. AIDS, Alzheimer's disease,
CC atherosclerosis, bacterial, fungal, protozoan or viral infections,
CC osteoarthritis, osteoporosis, cancer, cardiomyopathy, chronic and acute
CC inflammation, allergies, Crohn's disease, diabetes, graft versus host
CC disease, Parkinson's disease, multiple sclerosis, pain, psoriasis,
CC anxiety, depression, schizophrenia, dementia, mental retardation, memory
CC loss, epilepsy, asthma, tuberculosis, obesity, nausea, hypertension, or
CC hypotension, renal disorders, rheumatoid arthritis, trauma, ulcers, or
CC any other disorder in which GPCRs are involved. The antibodies may be
CC used in immunoassays and immunodiagnosis. ABZ42523 to ABZ42869 encode
CC GPCR proteins given in ABP81675 to ABP82018, which are used in the
CC exemplification of the present invention
XX
SQ Sequence 2456 BP; 519 A; 722 C; 571 G; 644 T; 0 U; 0 Other;
Query Match 100.0%; Score 42; DB 8; Length 2456;
Best Local Similarity 100.0%; Pred. No. 1.7e-07;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCCACGCTTGGCGAGCCCTTACAAGACTATGAAGTAAAG 42
Db 1271 GCCACGCTTGGCGAGCCCTTACAAGACTATGAAGTAAAG 1312
RESULT 25
ACC72695
ID ACC72695 standard; cDNA; 2456 BP.
XX
XX ACC72695;
AC

```





```
SQ Sequence 2456 BP; 519 A; 722 C; 571 G; 644 T; 0 U; 0 Other;
Query Match 100.0%; Score 42; DB 11; Length 2456;
Best Local Similarity 100.0%; Pred. No. 1.7e-07;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCCAGCGTTGGCGAGCCCTTACAAGACTATGAAGTAAAG 42
DB 1271 GCCCAGCGTTGGCGAGCCCTTACAAGACTATGAAGTAAAG 1312

RESULT 27
ADI28525
ID ADI28525 standard; cDNA; 2456 BP.
XX AC ADI28525;
XX DT 22-APR-2004 (first entry)
XX DE Human GPCR retinoic acid induced 3 (RAI-3) cDNA.
XX KW Retinoic acid induced 3; RAI-3; human; G-protein coupled receptor; GPCR;
XX KW antiinflammatory; immunosuppressive; cytostatic; cardiant; antiallergic;
XX KW broncholytic; gene therapy; gene; single nucleotide polymorphism; SNP;
XX KW chromosome 12p13-p12.3; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT variation replace(112,r)
FT FT /*tag= a
FT FT /label= RAI-3-s1
FT FT /note= "located in 5' untranslated region"
FT FT /standard_name= "Single nucleotide polymorphism"
FT CDS 254..1327
FT FT /*tag= b
FT FT /product= "Human RAI3"
FT FT replace(364,y)
FT FT /*tag= c
FT FT /label= RAI-3-s2
FT FT /note= "exon 1, silent (Ala/Ala)"
FT FT /standard_name= "Single nucleotide polymorphism"
FT FT replace(511,y)
FT FT /*tag= d
FT FT /label= RAI-3-s3
FT FT /note= "exon 2, silent (Ile/Ile)"
FT FT /standard_name= "Single nucleotide polymorphism"
FT FT replace(523,y)
FT FT /*tag= e
FT FT /label= RAI-3-s4
FT FT /note= "exon 2, silent (Asp/Asp)"
FT FT /standard_name= "Single nucleotide polymorphism"
FT FT replace(605,r)
FT FT /*tag= f
FT FT /label= RAI-3-s6
FT FT /note= "exon 2, missense (Ser/Gly)"
FT FT /standard_name= "Single nucleotide polymorphism"
FT FT replace(797,r)
FT FT /*tag= g
FT FT /label= RAI-3-s5
FT FT /note= "exon 2, missense (Thr/Ala)"
FT FT /standard_name= "Single nucleotide polymorphism"
FT FT replace(1111,r)
FT FT /*tag= h
FT FT /label= RAI-3-s8
FT FT /note= "silent (Pro/Pro)"
FT FT /standard_name= "Single nucleotide polymorphism"
FT FT replace(1173,r)
FT FT /*tag= i
FT FT /label= RAI-3-s9
FT FT /note= "missense (Gln/Arg)"
FT FT /standard_name= "Single nucleotide polymorphism"
XX
```

PN WO2004001060-A2.  
XX 31-DEC-2003.  
XX 20-JUN-2003; 2003WO-US019255.  
XX 20-JUN-2002; 2002US-0390850P.  
XX 29-AUG-2002; 2002US-0407006P.  
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.  
XX  
XX Whitney GS, Opitack G, Garulacan L, Ramanathan CS, McKinnon M;  
XX Bennett KL, Barber LB, Cacace A, Tsuchihashi Z;  
XX WPI; 2004-090973/09.  
XX P-PSDB; ADI28526.  
XX  
XX New nucleic acid molecule encoding a human G-protein coupled receptor,  
XX RAI3, useful for preventing, treating or ameliorating chronic obstructive  
XX pulmonary disease (COPD), COPD-like disorder, or the underlying symptoms  
XX of COPD.  
XX  
XX Claim 21; SEQ ID NO 18; 301pp; English.  
XX  
XX The present sequence is that of cDNA encoding a human G-protein coupled  
XX receptor (Class C, Group 5), denoted retinoic acid induced 3 (RAI-3).  
XX Proteomics methods were used to isolate cigarette smoke-inducible  
XX tyrosine phosphorylated proteins from airway epithelial cells. RAI-3 was  
XX identified as being tyrosine phosphorylated and/or as being  
XX associated/complexed with tyrosine phosphorylated proteins only in those  
XX cells that had been exposed to cigarette smoke. Since RAI-3 is primarily  
XX expressed in lung tissue, and since cigarette smoke is a major causative  
XX factor of chronic obstructive pulmonary disease (COPD), RAI-3 provides a  
XX novel cellular target for identifying modulators, e.g. agonists or  
XX antagonists, useful for the treatment and/or prevention of COPD and  
XX related disorders such as emphysema and chronic bronchitis. RAI-3  
XX modulators, e.g. agonists and antagonists, especially antisense  
XX compounds, can be used to treat COPD and/or its associated or diseases  
XX associated with regulation of NF-kB and/or its associated or interacting  
XX signaling molecules. Single nucleotide polymorphisms (SNPs) detected in  
XX the RAI-3 gene are useful for determining COPD association in  
XX individuals. RAI-3 nucleic acid molecules and polypeptides are useful for  
XX preventing, treating or ameliorating disorders related to aberrant GPCR  
XX signaling or cell cycle regulation, pulmonary disorders, inflammatory  
XX lung disorders, COPD, the underlying symptoms of COPD, COPD-related  
XX disorders or conditions, autoimmune disorders, disorders related to  
XX hyperimmune activity, inflammatory conditions, disorders related to  
XX aberrant acute phase responses, hypercongenital conditions, birth  
XX defects, necrotic lesions, wounds, organ transplant rejection, renal  
XX diseases, ischaemia-reperfusion injury, heart disorders, disorders  
XX related to aberrant signal transduction, proliferation disorders,  
XX cancers, HIV infection, asthma, cystic fibrosis, pulmonary fibrosis,  
XX ulcerative colitis, cerebral infarct, myocardial infarct, diabetic  
XX nephropathy, allergic rhinitis, Crohn's disease, atherosclerosis,  
XX rheumatoid arthritis, inflammatory/autoimmune disorders, glioblastoma,  
XX pulmonary small cell undifferentiated carcinoma, carcinoma of the breast,  
XX colon, lung, ovary, pancreas, prostate, non-Hodgkin's lymphoma, disorders  
XX associated with aberrant cell adhesion, I-CAM function and/or regulation,  
XX E-selectin function and/or regulation, or aberrant NF-kB function and/or  
XX regulation (all claimed).  
XX  
XX Sequence 2456 BP; 516 A; 719 C; 570 G; 643 T; 0 U; 8 Other;  
SQ

Query Match 100.0%; Score 42; DB 12; Length 2456;  
Best Local Similarity 100.0%; Pred. No. 1.7e-07;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCCAGCGTTGGCGAGCCCTTACAAGACTATGAAGTAAAG 42  
DB 1271 GCCCAGCGTTGGCGAGCCCTTACAAGACTATGAAGTAAAG 1312

RESULT 28

```

ADI28459
ID ADI28459 standard; cDNA; 2456 BP.
XX
AC ADI28459;
XX
DT 22-APR-2004 (first entry)
XX
DE Human GPCR retinoic acid induced 3 (RAI-3) cDNA.
XX
KW Retinoic acid induced 3; RAI-3; human; G-protein coupled receptor; GPCR;
KW antiinflammatory; immunosuppressive; cytostatic; cardiant; antiallergic;
KW broncholytic; gene therapy; gene; single nucleotide polymorphism; SNP;
KW chromosome 12p13-p12.3; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT variation replace(112,a)
FT *tag= a
FT /label= RAI-3-s1
FT /note= "located in 5' untranslated region"
FT /standard_name= "Single nucleotide polymorphism"
FT CDS 254..1327
FT *tag= b
FT /product= "Human RAI3"
FT variation replace(364,t)
FT *tag= c
FT /label= RAI-3-s2
FT /note= "exon 1, silent (Ala/Ala)"
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(511,t)
FT *tag= d
FT /label= RAI-3-s3
FT /note= "exon 2, silent (Ile/Ile)"
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(523,t)
FT *tag= e
FT /label= RAI-3-s4
FT /note= "exon 2, silent (Asp/Asp)"
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(605,g)
FT *tag= f
FT /label= RAI-3-s6
FT /note= "exon 2, missense (Ser/Gly)"
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(797,g)
FT *tag= g
FT /label= RAI-3-s5
FT /note= "exon 2, missense (Thr/Ala)"
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(1111,g)
FT *tag= h
FT /label= RAI-3-s8
FT /note= "silent (Pro/Pro)"
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(1173,g)
FT *tag= i
FT /label= RAI-3-s9
FT /note= "missense (Gln/Arg)"
FT /standard_name= "Single nucleotide polymorphism"
XX
PN WO2004001060-A2.
XX
PD 31-DEC-2003.
XX
PF 20-JUN-2003; 2003WO-US019255.
XX
PR 20-JUN-2002; 2002US-0390850P.
PR 29-AUG-2002; 2002US-0407006P.
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX Whitney GS, Opiteck G, Garulacan L, Ramanathan CS, McKinnon M;

```

```

Bennett KL, Barber LE, Cacace A, Tsuchihashi Z;
WPI; 2004-090973/09.
P-PSDB; ADI28460.
GENEANK; NM_003979.
New nucleic acid molecule encoding a human G-protein coupled receptor,
RAI3, useful for preventing, treating or ameliorating chronic obstructive
pulmonary disease (COPD), COPD-like disorder, or the underlying symptoms
of COPD.
Claim 1; SEQ ID NO 2; 301pp; English.
The present sequence is that of cDNA encoding a human G-protein coupled
receptor (Class C, Group 5), denoted retinoic acid induced 3 (RAI-3).
Proteomics methods were used to isolate cigarette smoke-inducible
tyrosine phosphorylated proteins from airway epithelial cells. RAI-3 was
identified as being tyrosine phosphorylated and/or as being
associated/complexed with tyrosine phosphorylated proteins only in those
cells that had been exposed to cigarette smoke. Since RAI-3 is primarily
expressed in lung tissue, and since cigarette smoke is a major causative
factor of chronic obstructive pulmonary disease (COPD), RAI-3 provides a
novel cellular target for identifying modulators, e.g. agonists or
antagonists, useful for the treatment and/or prevention of COPD and
related disorders such as emphysema and chronic bronchitis. RAI-3
modulators, e.g. agonists and antagonists, especially antisense
compounds, can be used to treat COPD and other disorders and diseases
associated with regulation of NF-kB and/or its associated or interacting
signaling molecules. Single nucleotide polymorphisms (SNPs) detected in
the RAI-3 gene are useful for determining COPD association in
individuals. RAI-3 nucleic acid molecules and polypeptides are useful for
preventing, treating or ameliorating disorders related to aberrant GPCR
signaling or cell cycle regulation, pulmonary disorders, inflammatory
lung disorders, COPD, the underlying symptoms of COPD, COPD-related
disorders or conditions, autoimmune disorders, disorders related to
hyperimmune activity, inflammatory conditions, disorders related to
aberrant acute phase responses, hypercongenital conditions, birth
defects, necrotic lesions, wounds, organ transplant rejection, renal
diseases, ischaemia-reperfusion injury, heart disorders, disorders
related to aberrant signal transduction, proliferation disorders,
cancers, HIV infection, asthma, cystic fibrosis, pulmonary fibrosis,
ulcerative colitis, cerebral infarct, myocardial infarct, diabetic
nephropathy, allergic rhinitis, Crohn's disease, atherosclerosis,
rheumatoid arthritis, inflammatory/autoimmune disorders, glioblastoma,
CC pneumonia small cell undifferentiated carcinoma, carcinoma of the breast,
CC colon, lung, ovary, pancreas, prostate, non-Hodgkin's lymphoma, disorders
CC associated with aberrant cell adhesion, I-CAM function and/or regulation,
CC E-selectin function and/or regulation, or aberrant NF-kB function and/or
CC regulation (all claimed).
SQ Sequence 2456 BP; 519 A; 722 C; 571 G; 644 T; 0 U; 0 Other;
Query Match 100.0%; Score 42; DB 12; Length 2456;
Best Local Similarity 100.0%; Pred. No. 1.7e-07;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCCACGCTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42
Db 1271 GCCACGCTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1312
RESULT 29
ADR48221
ID ADR48221 standard; cDNA; 2456 BP.
XX
AC ADR48221;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human retinoic acid induced 3 (RAI3) encoding cDNA SEQ:9.
XX
KW pancreatic cancer-associated transcript; pancreatic cancer; human;
KW cytostatic; gene therapy; protein therapy; retinoic acid induced 3; RAI3;

```

```

KW gene; ss.
XX Homo sapiens.
OS
XX
XX Location/Qualifiers
FH 254..1327
FT CDS /tag= a
FT /product= "retinoic acid induced 3 (RAI3)"
XX
XX WO2004074510-A1.
XX
XX 02-SEP-2004.
XX
XX 18-FEB-2004; 2004WO-AU000194.
XX
XX 18-FEB-2003; 2003AU-00900747.
XX
XX (GARV-) GARVAN INST MEDICAL RES.
XX
XX Biankin A, Segara D, Henshall S, Sutherland R;
XX
XX WPI; 2004-635591/61.
XX P-PSDB; ADR48222.
XX
XX Detecting pancreatic cancer-associated transcript in a biological sample,
XX useful for diagnosing or treating the disease, comprises contacting the
XX sample with a polynucleotide that selectively hybridizes to a specific
XX sequence.
XX
XX Claim 7; SEQ ID NO 9; 263pp; English.
XX
XX The present invention describes a method for detecting a pancreatic
XX cancer-associated transcript in a biological sample. The method comprises
XX contacting the biological sample with a polynucleotide that selectively
XX hybridizes to a sequence at least 80% identical to a sequence as shown in
XX any one of Tables 3 to 25 in the specification or having the GenBank
XX Accession Number AF279145. Also described: (1) diagnosing pancreatic
XX cancer in a human or animal subject being tested, determining the
XX likelihood that a subject having a pancreatic cancer will survive, or
XX determining the suitability of a subject having a pancreatic cancer for
XX surgical resection therapy; (2) detecting a pancreatic cancer-associated
XX polypeptide in a biological sample; (3) determining the likelihood that a
XX subject having a pancreatic cancer will survive; and (4) monitoring the
XX efficacy of a therapeutic treatment of pancreatic cancer. A pancreatic
XX cancer-associated transcript has cytostatic activity, and can be used in
XX gene and protein therapy. A pancreatic cancer-associated transcript
XX polynucleotide, a vector comprising the polynucleotide, an isolated
XX polypeptide or an antibody that binds to the isolated polypeptide can be
XX used for diagnosing or prognosing pancreatic cancer or for preparing a
XX medicament for the treatment of pancreatic cancer. The prognostic or
XX diagnostic methods are useful for the early detection of pancreatic
XX cancer or its metastases, and for monitoring the progress of disease such
XX as during remission or following surgery or chemotherapy. The present
XX sequence encodes human retinoic acid induced 3 (RAI3), which is used in
XX the exemplification of the present invention.
XX
XX Sequence 2456 BP; 519 A; 722 C; 571 G; 644 T; 0 U; 0 Other;
XX
XX Query Match 100.0%; Score 42; DB 13; Length 2456;
XX Best Local Similarity 100.0%; Pred. No. 1.7e-07;
XX Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 GCCCAGCGTTGGCCGAGCCCTTACAAGACTATGAAGTAAAG 42
XX |||||
XX 1271 GCCCAGCGTTGGCCGAGCCCTTACAAGACTATGAAGTAAAG 1312
XX
XX RESULT 30
XX ACN39388
XX ID ACN39388 standard; cDNA; 2456 BP.
XX
XX AC ACN39388;
XX

```

XX DT 18-NOV-2004 (first entry)  
XX DE Human breast tumour associated gene clone-02 SEQ ID NO:2.  
XX DE  
XX KW cancer; tumour; breast cancer; lung cancer; colon cancer; kidney cancer;  
KW antibody; cytostatic; gene therapy; human; breast tumour tissue;  
KW breast tissue; gene; ds.  
XX OS Homo sapiens.  
XX PN WO2004074506-A2.  
XX PD 02-SEP-2004.  
XX PF 30-JAN-2004; 2004WO-US003030.  
XX PR 13-FEB-2003; 2003US-0447900P.  
XX PA (MERC-) MERGEN LTD.  
XX PI Hu Q, Peng A, Liu B, Love JR, Hao X, Ren M, Sheng Z;  
XX WPI: 2004-635589/61.  
XX DR Diagnosing and treating a cancer, e.g. breast, lung, colon, or kidney  
XX cancer, comprises detecting and blocking the over expression of a gene of  
XX a protein found in breast, lung, colon, or kidney tissue.  
XX PS Claim 2; SEQ ID NO 2; 220pp; English.  
XX CC The present invention describes a method for diagnosing and treating a  
CC cancer, e.g. breast, lung, colon, or kidney cancer. The method comprises  
CC detecting and blocking the over expression of a gene of a protein found  
CC in breast, lung, colon, or kidney tissue. Also described is an antibody  
CC or binding portion of an antibody that specifically binds a protein found  
CC in breast tissue, lung tissue, colon tissue, or kidney tissue. The  
CC antibody has cytostatic activity, and can be used in gene therapy. The  
CC methods, antibodies, polynucleotides and polypeptides from the present  
CC invention are useful for detecting, diagnosing, preventing and treating  
CC cancer, e.g. breast, lung, colon, or kidney cancer. The present sequence  
CC represents a nucleotide sequence given in the present invention, which is  
CC over expressed in human breast tumour tissue. N.B. All 385 sequences  
CC referenced in this patent are detailed in the US provisional application  
CC SN 60/447,900, filed 02/13/2003. In this application only one sequence,  
CC representing the longest sequence of each of the 65 clones is listed in  
CC Appendix A (and given as SEQ ID NO:1 to 65 in the Sequence Listing).  
XX SQ Sequence 2456 BP; 517 A; 710 C; 567 G; 643 T; 0 U; 19 Other;  
Query Match 100.0%; Score 42; DB 13; Length 2456;  
Best Local Similarity 100.0%; Pred. No. 1.7e-07;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GCCCAGCGTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
DB 1271 GCCCAGCGTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1312  
RESULT 32  
ADU06126  
ID ADU06126 standard; DNA; 2456 BP.  
XX AC  
XX AC ADU06126;  
XX DT 27-JAN-2005 (first entry)  
XX DE Novel bronchial cancer-associated human gene SeqID350.  
XX KW bronchial cancer; cytostatic; tumour-associated protein;  
KW cancer detection; metastasis; tumour; gene; ds; human.  
XX OS Homo sapiens.

XX DE10316701-A1.  
XX PD 04-NOV-2004.  
XX PF 09-APR-2003; 2003DE-01016701.  
XX PR 09-APR-2003; 2003DE-01016701.  
XX PA (HINZ/) HINZMANN B.  
XX PA (HERM/) HERMANN K.  
XX PA (CAST/) HEIDEN CASTANOS-VELEZ E.  
XX PI Mennerich D, Bruemendorf T, Heiden E, Hermann K, Kinnemann H;  
PI Li X, Roepcke S, Staub E, Hinzmann B, Rosenthal A, Pilarsky C;  
XX WPI: 2004-786403/78.  
XX DR P-PSDB; ADU06613.  
XX PT New nucleic acid, and derived proteins, useful for diagnosis of bronchial  
XX cancer and in screening for therapeutic and diagnostic agents.  
XX PS Claim 1; SEQ ID NO 350; 1381pp; German.  
XX CC This invention relates to a novel isolated nucleic acid associated with  
CC bronchial cancer comprising 489 defined sequences given in the  
CC specification. The invention may be useful for the production of  
CC compounds with a cytostatic activity through the inhibition of expression  
CC or activity of tumour-associated proteins. The novel DNA sequences and  
CC the proteins/peptides encoded by them are used for detecting bronchial  
CC cancer or determining the risk of developing it and to screen for  
CC specific binding partners of the DNA or protein sequences, where the  
CC binding partners are potentially useful as agents for treating or  
CC diagnosing bronchial cancer. The DNA or protein sequences can also be  
CC used for prognosis, detection of metastases and for secondary treatment  
CC (of tumours that have been stabilised or are no longer detectable).  
CC Detecting abnormal expression of the DNA sequences provides early  
CC diagnosis of bronchial cancers. The present sequence is that of a novel  
CC bronchial cancer-associated human gene sequence of the invention.  
XX SQ Sequence 2456 BP; 519 A; 722 C; 571 G; 644 T; 0 U; 0 Other;  
Query Match 100.0%; Score 42; DB 13; Length 2456;  
Best Local Similarity 100.0%; Pred. No. 1.7e-07;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GCCCAGCGTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
DB 1271 GCCCAGCGTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1312  
RESULT 33  
AEA00088  
ID AEA00088 standard; cDNA; 2456 BP.  
XX AC AEA00088;  
XX DT 28-JUL-2005 (first entry)  
XX DE Human TAT115 cDNA SEQ ID NO:40.  
XX KW ss; gene; tumor-associated antigen; cytostatic; breast tumor;  
KW endocrine disease; gynecology and obstetrics; neoplasm; colon tumor;  
KW gastrointestinal disease; rectal tumor; endometroid carcinoma;  
KW genitourinary disease; renal tumor; lung tumor; respiratory disease;  
KW ovary tumor; skin tumor; liver tumor.  
XX OS Homo sapiens.  
XX PH Key Location/Qualifiers  
XX CDS 254..1327  
XX FT /\*tag= a  
XX FT /product= "TAT115"

```
XX US2005106644-A1.
XX
XX 19-MAY-2005.
XX
XX 08-SEP-2004; 2004US-00936626.
XX
XX 20-JUN-2001; 2001US-0299500P.
XX
XX 29-JUN-2001; 2001US-0301880P.
XX
XX 18-SEP-2001; 2001US-0323268P.
XX
XX 19-JUN-2002; 2002US-00177488.
XX
XX 26-MAR-2004; 2004US-0557116P.
XX
XX 04-AUG-2004; 2004US-0598899P.
XX
XX (GETH ) GENENTECH INC.
XX
XX Cairns B, Chen R, Frantz G, Hillan KJ, Koepfen H, Phillips HS;
XX Polakis P, Spencer SD, Smith V, Williams PM, Wu TD, Zhang Z;
XX Sliwkowski M;
XX
XX WPI; 2005-384304/39.
XX P-PSDB; AEA00166.
XX
XX Novel isolated antibody capable of binding to tumor-associated antigenic
XX target polypeptide, useful for treating cell proliferative disorder e.g.
XX cancer.
XX
XX Claim 1; SEQ ID NO 40; 337pp; English.
XX
XX The invention relates to a novel isolated antibody binding to a
XX polypeptide having at least 80% sequence identity to a polypeptide having
XX any one of 76 fully defined 182-910 amino acid tumor-associated antigenic
XX target polypeptide (TAT) sequences (AEA00127-AEA00202) given in the
XX specification, a polypeptide having any one of (AEA00127-AEA00202),
XX lacking its associated signal peptide, or an extracellular domain of a
XX polypeptide having any one of (AEA00127-AEA00202). The polypeptide is
XX encoded by the nucleotide sequence having any one of (AEA00049-AEA00126).
XX An antibody of the invention has cytostatic activity. The antibody is
XX useful for inhibiting growth of a cell expressing TAT188, which involves
XX contacting the cell with the antibody. The cell is a cancer cell chosen
XX from breast, colon, rectum, endometrium, kidney, lung, ovary, skin and
XX liver cell. The cancer cell is a mammalian cell, preferably a human cell.
XX The antibody is also useful for detecting the level of TAT188
XX polypeptide expressed in a test cell relative to a control cell, and for
XX detecting the level of TAT188 polypeptide or a polypeptide having at
XX least 80 % sequence identity to the TAT188 polypeptide sequence in a test
XX cell relative to a control cell. The antibody is useful for treating a
XX preventing a cell proliferative disorder associated with increased
XX expression or activity of a polypeptide having at least 80 % identity to
XX a TAT188 polypeptide sequence. The cell proliferative disorder is cancer.
XX The method of the invention is useful for inhibiting the growth of a
XX cancer cell. The present sequence encodes a polypeptide of the invention.
XX
XX Sequence 2456 BP; 519 A; 722 C; 571 G; 644 T; 0 U; 0 Other;
XX
XX Query Match 100.0%; Score 42; DB 14; Length 2456;
XX Best Local Similarity 100.0%; Pred. No. 1.7e-07;
XX Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 GCCCAGCTTGGCCGAGCCCTTACAAGACTATGAAGTAAAG 42
XX |||||||||||||||||||||||||||||||||||||||
XX Db 1271 GCCCAGCTTGGCCGAGCCCTTACAAGACTATGAAGTAAAG 1312
XX
XX RESULT 34
XX AEA00608
XX ID AEA00608 standard; cDNA; 2456 BP.
XX
XX AC AEA00608;
XX
XX 28-JUL-2005 (first entry)
XX
XX Human TAT115 cDNA sequence SeqID40.
XX
```

```
XX
XX antibody identification; tumor-associated antigen; cytostatic;
XX RNA interference; gene therapy; cell death; cancer; breast tumor;
XX colon tumor; rectal tumor; renal tumor; lung tumor; ovary tumor;
XX skin tumor; liver tumor; gene; ss; TAT115.
XX
XX Homo sapiens.
XX
XX US2005107595-A1.
XX
XX 19-MAY-2005.
XX
XX 10-SEP-2004; 2004US-00938061.
XX
XX 20-JUN-2001; 2001US-0299500P.
XX
XX 29-JUN-2001; 2001US-0301880P.
XX
XX 18-SEP-2001; 2001US-0323268P.
XX
XX 19-JUN-2002; 2002US-00177488.
XX
XX 26-MAR-2004; 2004US-0557116P.
XX
XX 04-AUG-2004; 2004US-0598899P.
XX
XX (GETH ) GENENTECH INC.
XX
XX Cairns B, Chen R, Frantz G, Hillan KJ, Koepfen H, Phillips HS;
XX Polakis P, Spencer SD, Smith V, Williams PM, Wu TD, Zhang Z;
XX Sakanaka C, Chundharapai A, Reed CJ;
XX
XX WPI; 2005-371577/38.
XX P-PSDB; AEA00686.
XX
XX Novel isolated antibody e.g. anti-E16 or anti-TAT112 antibody that binds
XX to tumor-associated antigenic target polypeptide, useful for diagnosing
XX or treating cancer.
XX
XX Claim 1; SEQ ID NO 40; 96pp; English.
XX
XX This invention relates to a novel isolated antibody, for example anti-E16
XX or anti-TAT112 antibody, that binds to a tumor-associated antigenic
XX target polypeptide (TAT) and that lacks an associated signal peptide
XX sequence. The invention may be useful for the development of compounds
XX with a cytostatic activity acting as antagonists of the TAT188
XX polypeptide or RNA interference whilst the disclosed sequences may be
XX useful for gene therapy. The invention is useful for inducing the death
XX of a cell (such as a cancer cell chosen from breast, colon, rectum,
XX endometrium, kidney, lung, ovary, skin and liver) to which it binds,
XX inhibiting proliferation or promoting cell death of a cell expressing
XX TAT188. In addition, the invention may be useful for detecting the level
XX of TAT188 polypeptide in a test cell relative to a control cell, or
XX treating or preventing a cell proliferative disorder associated with
XX increased expression of TAT188. The novel antibody of the invention is
XX useful for inhibiting the growth of a cancer cell and may be useful for
XX diagnosing or treating cancer. The present sequence is that of the human
XX TAT115 cDNA which encodes a protein against which an antibody of the
XX invention may be targeted.
XX
XX Sequence 2456 BP; 519 A; 722 C; 571 G; 644 T; 0 U; 0 Other;
XX
XX Query Match 100.0%; Score 42; DB 14; Length 2456;
XX Best Local Similarity 100.0%; Pred. No. 1.7e-07;
XX Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 GCCCAGCTTGGCCGAGCCCTTACAAGACTATGAAGTAAAG 42
XX |||||||||||||||||||||||||||||||||||||||
XX Db 1271 GCCCAGCTTGGCCGAGCCCTTACAAGACTATGAAGTAAAG 1312
XX
XX RESULT 35
XX AED47480
XX ID AED47480 standard; DNA; 2456 BP.
XX
XX AC AED47480;
XX
XX 15-DEC-2005 (first entry)
XX
```

XX DE Retinoic acid-inducible G-protein coupled receptor 3 gene, SEQ ID 2.  
XX KW RNA Interference; gene silencing;  
XX KW retinoic acid-inducible G-protein coupled receptor 3; breast tumor;  
XX KW cytostatic; cell growth; gene; ds.  
XX OS Homo sapiens.  
XX  
XX FH Location/Qualifiers  
XX CDS 254..1327  
XX FT /\*tag= a  
XX FT /product= "Retinoic acid-inducible G-protein coupled  
XX FT receptor 3"  
XX FT misc\_feature 338..403  
XX FT /\*tag= b  
XX FT /label= Transmembrane domain  
XX FT misc\_feature 452..514  
XX FT /\*tag= c  
XX FT /label= Transmembrane domain  
XX FT misc\_feature 497..1057  
XX FT /\*tag= d  
XX FT /label= 7 transmembrane receptor  
XX FT misc\_feature 548..601  
XX FT /\*tag= e  
XX FT /label= Transmembrane domain  
XX FT misc\_feature 656..718  
XX FT /\*tag= f  
XX FT /label= Transmembrane domain  
XX FT misc\_feature 776..832  
XX FT /\*tag= g  
XX FT /label= Transmembrane domain  
XX FT misc\_feature 887..946  
XX FT /\*tag= h  
XX FT /label= Transmembrane domain  
XX FT polyA\_signal 2403..2408  
XX FT /\*tag= i  
XX FT polyA\_site 2442..2456  
XX FT /\*tag= j  
XX JP2005287456-A.  
XX  
XX PD 20-OCT-2005.  
XX PF 02-APR-2004; 2004JP-00110232.  
XX PR 02-APR-2004; 2004JP-00110232.  
XX PA (NIPK ) NIPPON KAYAKU KK.  
XX PA (UYNI-) UNIV NIPPON IKA.  
XX PI Nagahata T, Emi M;  
XX  
XX DR WPI: 2005-737982/76.  
XX DR P-PSDB; AED47479.  
XX  
XX PT Substance that suppresses expression of retinoic acid-inducible G-protein  
XX PT coupled receptor-3 RA13 gene or nucleic acid, useful for suppressing cell  
XX PT growth, and as pharmaceutical or therapeutic agent of breast cancer.  
XX  
XX PS Claim 1; SEQ ID NO 2; 22pp; Japanese.  
XX  
XX CC The invention relates to a novel substance which suppresses the  
XX CC expression of a retinoic acid-inducible G-protein coupled receptor (RA1)3  
XX CC gene. The RA13 gene suppressing agent is useful for suppressing cell  
XX CC growth, and as a pharmaceutical or therapeutic agent of breast cancer,  
XX CC and is also useful in identifying a cell growth inhibitory substance. The  
XX CC RA13 gene suppressing agent enables the identification of a cell growth  
XX CC inhibitory substance. This polynucleotide sequence represents the  
XX CC retinoic acid-inducible G-protein coupled receptor 3 gene of the  
XX CC invention.  
XX  
XX SQ Sequence 2456 BP; 519 A; 722 C; 571 G; 644 T; 0 U; 0 Other;

Query Match 100.0%; Score 42; DB 14; Length 2456;  
Best Local Similarity 100.0%; Pred. No. 1.7e-07;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GCCCAGCGTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
DB 1271 GCCCAGCGTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1312

RESULT 36  
ABQ54954  
ID ABQ54954 standard; cDNA; 2593 BP.  
XX AC ABQ54954;  
XX DT 22-AUG-2002 (first entry)  
XX DE Human ovarian antigen HSKJC61 cDNA, SEQ ID NO:834.  
XX KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;  
XX KW ovarian cancer; breast cancer; tumour; reproductive system disorder;  
XX KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;  
XX KW PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;  
XX KW inflammatory condition; immune disorder; blood disorder;  
XX KW cardiovascular disorder; respiratory disorder; neurological disorder;  
XX KW gastrointestinal disorder; urinary system disorder; drug screening;  
XX KW gene therapy; chromosome mapping; forensic analysis;  
XX KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;  
XX KW antiinflammatory; gynaecological; reproductive; gene; ss.  
XX  
XX OS Homo sapiens.  
XX PN WO200200677-A1.  
XX PD 03-JAN-2002.  
XX PF 07-JUN-2001; 2001WO-US018569.  
XX PR 07-JUN-2000; 2000US-0209467P.  
XX PA (HUMA-) HUMAN GENOME SCI INC.  
XX PI Birse CE, Rosen CA;  
XX DR WPI: 2002-147878/19.  
XX DR P-PSDB; ABP41877.  
XX  
XX PT Isolated nucleic acid molecules encoding novel ovarian polypeptides,  
XX PT useful in the prevention, treatment and diagnosis of cancer (e.g. ovarian  
XX PT cancer), immune disorders, cardiovascular disorders and neurological  
XX PT diseases.  
XX  
XX PS Claim 1; SEQ ID NO 834; 2922pp; English.  
XX  
XX CC The invention relates to 2175 novel human ovarian antigens (ABP41054-  
XX CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also  
XX CC encompasses polypeptides 90% identical and polynucleotides 95% identical  
XX CC to the sequences of the invention. The invention additionally relates to  
XX CC recombinant vectors and host cells comprising human ovarian antigen  
XX CC polynucleotides, antibodies against human ovarian antigens, and the use  
XX CC of ovarian antigen polynucleotides and polypeptides in diagnosing,  
XX CC treating, prognosing or preventing various ovary and/or breast-related  
XX CC disorders. Such conditions include ovarian cancer and breast cancer, and  
XX CC metastatic tumours of ovarian or breast origin, reproductive system  
XX CC disorders (e.g., infertility, disorders of pregnancy, anovulation,  
XX CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhea), endocrine  
XX CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic  
XX CC vaginitis), immune disorders (e.g., congenital and acquired  
XX CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),  
XX CC blood-related disorders (e.g., anaemia), cardiovascular disorders,  
XX CC respiratory disorders, neurological disorders, gastrointestinal disorders

CC and urinary system disorders. Ovarian antigen polypeptides and  
CC polynucleotides may also be used in screening for compounds which  
CC modulate ovarian antigen expression or activity. The polynucleotides may  
CC further be used for gene therapy, chromosome mapping, in the  
CC identification of individuals and in forensic analysis, and the  
CC polypeptides may be used as food additives or to prepare antibodies  
CC useful in disease diagnosis, drug targeting and phenotyping. The present  
CC sequence represents cDNA encoding a human ovarian antigen of the  
CC invention. Note: The sequence data for this patent did not form part of  
CC the printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 2593 BP; 591 A; 730 C; 605 G; 667 T; 0 U; 0 Other;

Query Match 100.0%; Score 42; DB 6; Length 2593;  
Best Local Similarity 100.0%; Pred. No. 1.8e-07;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
Db 1383 GCCCAGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 1424

RESULT 37  
AAH72766  
ID AAH72766 standard; cDNA; 3371 BP.

XX AC AAH72766;

XX DT 19-SEP-2001 (first entry)

XX DE Human cervical cancer marker nucleic acid 4040.

XX KW Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.

XX OS Homo sapiens.

XX PN WO200142467-A2.

XX PD 14-JUN-2001.

XX PF 08-DEC-2000; 2000WO-US033312.

XX PR 08-DEC-1999; 99US-0169681P.

XX PR 21-DEC-1999; 99US-0171350P.

XX PR 14-MAR-2000; 2000US-0189315P.

XX PR 12-MAY-2000; 2000US-0203791P.

XX PR 09-JUN-2000; 2000US-0210600P.

XX PR 21-JUL-2000; 2000US-0220114P.

XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX PI Schlegel R, Deeds J, Berger A, Zhao X;

XX PS WPI; 2001-375006/39.

XX PT New isolated nucleic acid for diagnosing and treating cervical cancer and  
XX for assessing and detecting compounds for treating the cancer.

XX PS Claim 1; Page 840; 1051pp; English.

XX CC The invention relates to novel genes (AAH68727-AAH73383) associated with  
XX cervical cancer with cytostatic activity. The nucleic acids and encoded  
XX polypeptides are useful: to assess if a patient is afflicted with  
XX cervical cancer or has a pre-malignant condition; to monitor the  
XX progression of cervical cancer or a premalignant condition in a patient;  
XX and to select and/or assess the efficacy of a compound or therapy for  
XX inhibiting cervical cancer in a patient. The nucleic acids may also be  
XX useful for gene therapy

XX SQ Sequence 3371 BP; 741 A; 902 C; 779 G; 946 T; 0 U; 3 Other;

Query Match 100.0%; Score 42; DB 4; Length 3371;

Best Local Similarity 100.0%; Pred. No. 1.9e-07;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
Db 1184 GCCCAGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 1225

RESULT 38

ACN89274  
ID ACN89274 standard; DNA; 4239 BP.

XX AC ACN89274;

XX DT 02-DEC-2004 (first entry)

XX DE Breast cancer related marker, seq id 10424.

XX KW Cancer; breast; tumour; cytostatic; marker; detection; therapy; ds.

XX OS Homo sapiens.

XX PN US2003099974-A1.

XX PD 29-MAY-2003.

XX PF 18-JUL-2002; 2002US-00198846.

XX PR 18-JUL-2001; 2001US-0306220P.

XX PA (MILL-) MILLENNIUM PHARM INC.

XX PI Lillie J, Xu Y, Wang Y, Steinmann K;

XX DR WPI; 2003-787014/74.

XX PT Novel isolated polypeptide associated with breast cancer, useful for  
XX detecting presence of polypeptide in sample, as a marker for breast  
XX cancer.

XX PS Disclosure; SEQ ID NO 10424; 36pp; English.

XX CC The invention relates to an isolated polypeptide (I) associated with  
XX breast cancer which is encoded by a nucleic acid molecule comprising a  
XX nucleotide sequence (S1). Further disclosed is an antibody that binds to  
XX the polypeptide of the invention. The activity of the polypeptide of the  
XX invention may be described as cytostatic. The antibody is useful for  
XX detecting the presence of (I) in a sample. Nucleic acid molecules of the  
XX invention are useful in the detection of breast tumours. (I) is useful as  
XX a marker for breast cancer and in breast cancer therapy. Sequences given  
XX in records ACN78851-ACN92934 represent nucleic acid markers associated  
XX with breast cancer. Note: The sequence listing does not form part of the  
XX specification but may be obtained in electronic format from the USPTO web  
XX site at seqdata.uspto.gov/sequence.html?DocID=20030099974

XX SQ Sequence 4239 BP; 859 A; 1161 C; 1075 G; 1141 T; 0 U; 3 Other;

Query Match 100.0%; Score 42; DB 11; Length 4239;

Best Local Similarity 100.0%; Pred. No. 2e-07;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
Db 1679 GCCCAGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 1720

RESULT 39

ADX97494  
ID ADX97494 standard; DNA; 6730 BP.

XX AC ADX97494;

XX DT 21-APR-2005 (first entry)

```
XX DE Pancreatic cancer associated human protein encoding DNA, SEQ ID 42.
XX KW pancreas tumor; cytostatic; gene; ds.
XX OS Homo sapiens.
XX PN EP1471075-A2.
XX PD 27-OCT-2004.
XX PF 31-MAR-2004; 2004EP-00090124.
XX PR 31-MAR-2003; 2003DE-01015834.
XX PA (HINZ/) HINZMANN B.
XX PA (ROSE/) ROSENTHAL A.
XX PA (PILA/) PILARSKY C.
XX PA (DAHL/) DAHL E.
XX PA (SPEC/) SPECHT T.
XX PA (LICH/) LICHTNER R.
XX PI Rosenthal A, Pilarsky C, Dahl E, Specht T, Bruemendorf T;
XX PI Lichtner R, Staub E, Roepcke S, Li X;
XX DR WPI; 2004-768082/76.
XX DR P-PSDB; ADX97565.
XX PT New nucleic acid differentially expressed in pancreatic tumor tissue, for
XX PT use as diagnostic agents and in screening for therapeutic agents.
XX PS Claim 1; SEQ ID NO 42; 28pp; German.
XX CC The invention relates to a novel human nucleic acid sequence of the
XX CC pancreas and its encoded protein. The invention further comprises:
XX CC proteins and peptides, preferably isolated, that contain a sequence
XX CC encoded by the novel nucleic acid; and methods for diagnosis and
XX CC treatment of pancreatic cancer, using a substance that inhibits or binds
XX CC to the protein or its DNA, including: an antisense oligonucleotide, short
XX CC interfering RNA or ribozyme directed against the pancreatic protein, an
XX CC organic molecule, particularly having a molecular weight below 5000,
XX CC especially 300, that binds to the pancreatic DNA, an aptamer or
XX CC (monoclonal) antibody, preferably human or humanized, that binds to the
XX CC pancreatic DNA, or an anti-idiotypic antibody raised against the
XX CC monoclonal antibody, any of which may be derivatized with a reporter
XX CC group, cytotoxic compound, immunostimulant and/or radioisotope. The novel
XX CC human pancreatic proteins and their encoding DNA have cytostatic
XX CC activity. The novel sequences are useful for inhibiting transcription
XX CC and/or expression of genes and proteins associated with pancreatic
XX CC cancer. This polynucleotide sequence represents the DNA encoding one of
XX CC the novel human pancreatic proteins of the invention. Note: This sequence
XX CC is not shown in the specification, it has been electronically downloaded
XX CC from a DVD-rom provided with this specification by the European Patent
XX CC Office.
XX SQ Sequence 6730 BP; 1459 A; 1763 C; 1555 G; 1943 T; 0 U; 0 Other;
Query Match 100.0%; Score 42; DB 13; Length 6730;
Best Local Similarity 100.0%; Pred. No. 2.2e-07;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCCCAGCGTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42
DB 1271 GCCCAGCGTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1312
RESULT 40
ADF70573
ID ADF70573 standard; DNA; 1788 BP.
XX AC ADF70573;
XX AC ADF70573;
XX DT 12-FEB-2004 (first entry)
```

```
XX DE Orphan receptor ligand-related human protein gene SeqID196.
XX KW ligand; orphan receptor protein; fusion protein; fluorescent protein;
XX KW cell expression; green fluorescent protein; GFP; GFP-I; wild-type GFP;
XX KW GFPuv; Enhanced GFP; EGFP; human; gene; ds.
XX OS Homo sapiens.
XX PN WO2003071272-A1.
XX PD 28-AUG-2003.
XX PF 21-FEB-2003; 2003WO-JP001901.
XX PR 22-FEB-2002; 2002JP-00045728.
XX PR 23-JUL-2002; 2002JP-00213949.
XX PR 11-OCT-2002; 2002JP-00298237.
XX XX (TAKE ) TAKEDA CHEM IND LTD.
XX PI Hinuma S, Fujii R, Ogi K, Komatsu H, Kawamata Y, Hosoya M;
XX DR WPI; 2003-697654/66.
XX DR P-PSDB; ADF70471.
XX PT Transformation of cells with a fusion protein of an orphan receptor
XX PT protein with a fluorescent protein useful for identification of ligands
XX PT to the orphan receptor.
XX PS Example 4; SEQ ID NO 196; 594pp; Japanese.
XX CC This invention relates to a novel method of identifying ligands to an
XX CC orphan receptor protein which comprises transforming cells with DNA
XX CC encoding a fusion protein of the orphan receptor with a fluorescent
XX CC protein, so that the fusion protein is expressed in the cells (or cell
XX CC membranes isolated from them) and contacting the cells with the potential
XX CC ligand to be tested. A suitable fluorescent protein for incorporation in
XX CC the fusion protein is green fluorescent protein (GFP), for example GFP-I,
XX CC wild-type GFP, GFPuv or Enhanced GFP (EGFP). The method is useful for the
XX CC identification of ligands binding to an orphan receptor protein.
XX SQ Sequence 1788 BP; 444 A; 473 C; 412 G; 459 T; 0 U; 0 Other;
Query Match 96.2%; Score 40.4; DB 10; Length 1788;
Best Local Similarity 97.6%; Pred. No. 7.8e-07;
Matches 41; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GCCCAGCGTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42
DB 1018 GCCCAGCGTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1059
RESULT 41
AAH08354
ID AAH08354 standard; cDNA; 680 BP.
XX AC AAH08354;
XX DT 26-JUN-2001 (first entry)
XX DE Human cDNA clone (5'-primer) SEQ ID NO:5189.
XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX OS Homo sapiens.
XX PN EP1074617-A2.
XX PD 07-FEB-2001.
XX PF 28-JUL-2000; 2000EP-00116126.
```



```
PR 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT length cDNAs defined in the specification, and for the detection and/or
PT diagnosis of the abnormality of the proteins encoded by the full-length
PT cDNAs.
XX
XX Claim 1; SEQ ID NO 5189; 2537pp + Sequence Listing; English.
XX
XX The present invention describes primer sets for synthesizing 5602 full-
CC length cDNAs defined in the specification. Where a primer set comprises:
CC (a) an oligo-dT primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC polynucleotides, all of which are used in the exemplification of the
CC present invention
XX
XX Sequence 680 BP; 143 A; 194 C; 151 G; 184 T; 0 U; 8 Other;
SQ
Query Match 80.0%; Score 33.6; DB 4; Length 680;
Best Local Similarity 87.8%; Pred. No. 0.0005;
Matches 36; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 GCCCAGCTTGGCCGAGCCCTTACAAGACTATGTAAGTAA 41
DB 633 GCCCAGCTTGGCCCAACCTTACAAGACTATGTAAGTAA 673
RESULT 42
ADI28530
ID ADI28530 standard; DNA; 42 BP.
XX
XX ADI28530;
XX
XX 22-APR-2004 (first entry)
XX
XX Human GPCR retinoic acid induced 3 (RAI-3) peptide DNA.
XX
XX Retinoic acid induced 3; RAI-3; human; G-protein coupled receptor; GPCR;
KW antiinflammatory; immunosuppressive; cytostatic; cardiac; antiallergic;
KW broncholytic; gene therapy; ss.
XX
XX Homo sapiens.
OS Synthetic.
XX
XX WO2004001060-A2.
XX
```

```
PD 31-DEC-2003.
XX
XX 20-JUN-2003; 2003WO-US019255.
XX
XX 20-JUN-2002; 2002US-0390850P.
PR 29-AUG-2002; 2002US-0407006P.
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
XX Whitney GS, Opitack G, Garulacan L, Ramanathan CS, McKinnon M;
PI Bennett KL, Barber LE, Cacace A, Tsuchihashi Z;
XX WPI; 2004-090973/09.
XX
XX New nucleic acid molecule encoding a human G-protein coupled receptor,
PT RAI3, useful for preventing, treating or ameliorating chronic obstructive
PT pulmonary disease (COPD), COPD-like disorder, or the underlying symptoms
PT of COPD.
XX
XX Disclosure; SEQ ID NO 33; 301pp; English.
XX
XX The present sequence is that of DNA encoding a peptide corresponding to
CC amino acids 340-353 of a human G-protein coupled receptor, denoted
CC retinoic acid induced 3 (RAI-3) protein ADI28460. The peptide was
CC identified in studies of RAI-3 phosphorylation in response to cigarette
CC smoke. Primers based on this sequence can be used in a PCR to identify
CC individuals at risk for developing chronic obstructive pulmonary disease
CC (COPD). RAI-3 is tyrosine phosphorylated and/or is associated/complexed
CC with tyrosine phosphorylated proteins only in cells that have been
CC exposed to cigarette smoke. Since RAI-3 is primarily expressed in lung
CC tissue, and since cigarette smoke is a major causative factor of COPD,
CC RAI-3 provides a novel cellular target for identifying modulators, e.g.
CC agonists or antagonists, useful for the treatment and/or prevention of
CC COPD and related disorders such as emphysema and chronic bronchitis. RAI-
CC 3 modulators, e.g. agonists and antagonists, especially antisense
CC compounds, can be used to treat COPD and other disorders or diseases
CC associated with regulation of NF-kB and/or its associated or interacting
CC signaling molecules. Single nucleotide polymorphisms (SNPs) detected in
CC the RAI-3 gene are useful for determining COPD association in
CC individuals. RAI-3 nucleic acid molecules and polypeptides are useful for
CC preventing, treating or ameliorating disorders related to aberrant GPCR
CC signaling or cell cycle regulation, pulmonary disorders, inflammatory
CC lung disorders, COPD, the underlying symptoms of COPD, COPD-related
CC disorders or conditions, autoimmune disorders, disorders related to
CC hyperimmune activity, inflammatory conditions, disorders related to
CC aberrant acute phase responses, hypercongenital conditions, birth
CC defects, necrotic lesions, wounds, organ transplant rejection, renal
CC diseases, ischaemia-reperfusion injury, heart disorders, disorders
CC related to aberrant signal transduction, proliferation disorders,
CC cancers, HIV infection, asthma, cystic fibrosis, pulmonary fibrosis,
CC ulcerative colitis, cerebral infarct, myocardial infarct, diabetic
CC nephropathy, allergic rhinitis, Crohn's disease, atherosclerosis,
CC rheumatoid arthritis, inflammatory/autoimmune disorders, glioblastoma,
CC pulmonary small cell undifferentiated carcinoma, carcinoma of the breast,
CC colon, lung, ovary, pancreas, prostate, non-Hodgkin's lymphoma, disorders
CC associated with aberrant cell adhesion, I-CAM function and/or regulation,
CC E-selectin function and/or regulation, or aberrant NF-kB function and/or
CC regulation (all claimed).
XX
XX Sequence 42 BP; 9 A; 8 C; 7 G; 5 T; 0 U; 13 Other;
SQ
Query Match 71.4%; Score 30; DB 12; Length 42;
Best Local Similarity 64.3%; Pred. No. 0.0092;
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;
QY 1 GCCCAGCTTGGCCGAGCCCTTACAAGACTATGTAAGTAA 42
DB 1 GCNCAYGCTGGCCCTCNCCTTAYARGAYTAYARGTNAAR 42
RESULT 43
ADI28527
ID ADI28527 standard; DNA; 40 BP.
XX
```



Db 330 CACGCTGGCGAAGCCTTTTCAACGACTTTGAAG 363

## RESULT 45

ADV92553  
ID ADV92553 standard; DNA; 525 BP.

XX AC ADV92553;

XX DT 10-MAR-2005 (first entry)

XX DE Japanese encephalitis virus (JEV) M15EB protein encoding DNA.

XX DE gene; ds; Immune stimulation; virucide; vaccine; envelope protein;  
KW membrane protein; immunostimulant; flavivirus infection.

XX OS Japanese encephalitis virus.

XX FH Key Location/Qualifiers

XX CDS 1..525

FT /tag= b

FT /product= "Japanese encephalitis virus M15EB protein"

FT /transl\_except= (pos:1..5, aa:Val)

FT /partial

FT /note= "No start and stop codons"

FT misc\_feature 1..45

FT /tag= a

FT /note= "JEV M carboxyl-terminal peptide encoding DNA

FT fragment"

FT /note= "This fragment is stated as signal peptide

FT encoding DNA fragment"

FT misc\_feature 46..525

FT /tag= c

FT /note= "JEV EB carboxyl-terminal protein encoding DNA"

XX US2004254365-A1.

XX PN

XX PD 16-DEC-2004.

XX PP 11-JUN-2003; 2003US-00459155.

XX PR 11-JUN-2003; 2003US-00459155.

XX PA (LEES/) LEE S T.

XX PI Lee ST;

XX PP 11-JUN-2003; 2003US-00459155.

XX PR 11-JUN-2003; 2003US-00459155.

XX PA (LEES/) LEE S T.

XX PI Lee ST;

XX PP 11-JUN-2003; 2003US-00459155.

XX PR 11-JUN-2003; 2003US-00459155.

XX PA (LEES/) LEE S T.

XX PI Lee ST;

XX PP 11-JUN-2003; 2003US-00459155.

XX PR 11-JUN-2003; 2003US-00459155.

XX PA (LEES/) LEE S T.

XX PI Lee ST;

XX PP 11-JUN-2003; 2003US-00459155.

XX PR 11-JUN-2003; 2003US-00459155.

XX PA (LEES/) LEE S T.

XX PI Lee ST;

XX PP 11-JUN-2003; 2003US-00459155.

XX PR 11-JUN-2003; 2003US-00459155.

XX PA (LEES/) LEE S T.

XX PI Lee ST;

XX PP 11-JUN-2003; 2003US-00459155.

XX PR 11-JUN-2003; 2003US-00459155.

XX PA (LEES/) LEE S T.

XX PI Lee ST;

XX PP 11-JUN-2003; 2003US-00459155.

XX PR 11-JUN-2003; 2003US-00459155.

XX PA (LEES/) LEE S T.

XX PI Lee ST;

XX PP 11-JUN-2003; 2003US-00459155.

XX PR 11-JUN-2003; 2003US-00459155.

QY 4 CACGCTTGGCGAGCCCTTACAAAGACTATGAAG 37  
Db 375 CACGCTGGGCAAGCCTTTTCAACGACTTTGAAG 408

## RESULT 46

AAA29595  
ID AAA29595 standard; cDNA; 1500 BP.

XX AC AAA29595;

XX DT 09-AUG-2000 (first entry)

XX DE Japanese encephalitis virus protein JMSV001 encoding cDNA.

XX DE Japanese encephalitis virus; infection; immunogen; vaccine; diagnosis;  
KW viral particle; ThCAMAR67/93; JMSV001; antiviral; immunomodulatory; ss.

XX OS Japanese encephalitis virus.

XX PN WO200020565-A1.

XX PD 13-APR-2000.

XX PF 02-JUN-1999; 99WO-JP002931.

XX PR 05-OCT-1998; 98JP-00319762.

XX PA (OSAU ) UNIV OSAKA.

XX PI Ishikawa T, Yoshii H, Onishi T, Imagawa T, Ishibashi M;

XX DR WPI; 2000-303759/26.

XX P-PSDB; AAY82692.

XX Novel inactivated viral particles, useful as vaccines against and in the  
PT diagnosis of infection with Japanese encephalitis viruses, is prepared  
PT from an infective cell culture of a Japanese encephalitis virus.  
XX Disclosure; Page 33-36; 42pp; Japanese.

XX The present invention describes inactivated immunogenic viral particles  
CC (I) prepared from an infective cell culture of a Japanese encephalitis  
CC virus. (I) are useful for treatment, diagnosis and as vaccines against  
CC Japanese encephalitis virus infection. The present sequence encodes a  
CC Japanese encephalitis virus protein, designated JMSV001, which is given  
CC in the exemplification of the present invention

XX Sequence 1500 BP; 393 A; 362 C; 411 G; 334 T; 0 U; 0 Other;

Query Match 54.3%; Score 22.8; DB 3; Length 1500;

Best Local Similarity 79.4%; Pred. No. 24;

Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 CACGCTTGGCGAGCCCTTACAAAGACTATGAAG 37

Db 1203 CACGCTGGGCAAGCCTTTTCAACGACTTTGAAG 1236

## RESULT 47

ADV92550  
ID ADV92550 standard; DNA; 1500 BP.

XX AC ADV92550;

XX DT 10-MAR-2005 (first entry)

XX DE Japanese encephalitis virus (JEV) E gene.

XX gene; ds; Immune stimulation; virucide; vaccine; envelope protein;  
KW membrane protein; immunostimulant; flavivirus infection.

Query Match 54.3%; Score 22.8; DB 14; Length 525;

Best Local Similarity 79.4%; Pred. No. 19;

Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;



```

FH key Location/Qualifiers
FT sig_peptide 10..78
FT CDS /*tag= a
FT 79..579
FT /*tag= b
FT /product= "Japanese encephalitis virus prM protein"
FT mat_peptide 355..579
FT /*tag= c
FT /product= "Japanese encephalitis virus M protein"
FT CDS 580..2079
FT /*tag= d
FT /product= "Japanese encephalitis virus E protein"
FT JP2004065118-A.
FT 04-MAR-2004.
XX
XX
XX 07-AUG-2002; 2002JP-00229597.
XX
XX 07-AUG-2002; 2002JP-00229597.
XX
XX (HAND-) ZH HANDAI BISEIBUTSUBYO KENKYUKAI.
XX PA (KOKU-) KOKURITSU YOBO EISEI KENKYUSHO.
XX
XX WPI: 2004-209123/20.
XX P-PSDB; ADM97134, ADM97135, ADM97136, ADM97137.
XX
XX Novel Japanese-encephalitis-virus antigen comprising Japanese-
XX encephalitis-virus-like particle consisting of M and E protein, useful as
XX Japanese-encephalitis vaccine.
XX
XX Claim 3; SEQ ID NO 4; 45pp; Japanese.
XX
XX The invention relates to a novel Japanese-encephalitis-virus antigen
XX comprising a Japanese-encephalitis-virus-like particle, which includes
XX the M protein and E protein of a Japanese encephalitis virus, and which
XX does not contain RNA in the particle, and shows haemagglutination
XX activity. The invention further comprises: a transformed cell obtained by
XX providing cDNA encoding the E protein and prM protein prepared from the
XX genome RNA of a Japanese-encephalitis virus, preparing a recombinant DNA
XX by integrating the cDNA in an expression vector which contains a beta-
XX actin promoter and which can be expressed, transforming an animal cell by
XX the recombinant DNA, selecting the transformed cell from a parent cell
XX and making the transformed cell to acclimatize in serum free medium;
XX Japanese-encephalitis-vaccine having immunity for the novel antigen and
XX containing the antigen as an active ingredient; and a diagnostic agent
XX which contains the antigen as an active ingredient. The Japanese-
XX encephalitis-virus antigen is useful for carrying out the preparation of
XX a Japanese-encephalitis vaccine and a diagnostic agent. This
XX polynucleotide sequence represents the Japanese-encephalitis-virus
XX encoding DNA of the invention.
XX
XX Sequence 2088 BP; 555 A; 497 C; 574 G; 462 T; 0 U; 0 Other;
XX
XX Query Match 54.3%; Score 22.8; DB 12; Length 2088;
XX Best Local Similarity 79.4%; Pred. No. 25;
XX Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
XX
XX 4 CACGCTTGCGCGAGCCCTTACAAGACTATGAAG 37
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 1782 CACGCTGGGCAAGCCCTTTCAACGACTTTGAAG 1815
XX
XX RESULT 50
XX AAQ22767
XX ID AAQ22767 standard; DNA; 4512 BP.
XX
XX AAQ22767;
XX
XX 12-AUG-1992 (first entry)
XX
XX JEV Nakayama strain prM, E, NS1, NS2A, NS2B and C coding regions.
XX

```

```

KW Japanese Encephalitis Virus; vaccinia virus donor; plasmid pDr20; ss.
XX
XX Japanese encephalitis virus.
XX
XX WO9203545-A.
XX
XX 05-MAR-1992.
XX
XX 05-AUG-1991; 91WO-US005816.
XX
XX 15-AUG-1990; 90US-00567960.
XX
XX 06-JUN-1991; 91US-00711429.
XX
XX 13-JUN-1991; 91US-00714687.
XX
XX 17-JUL-1991; 91US-00729800.
XX
XX 05-AUG-1991; 91WO-U0005816.
XX
XX (VIRO-) VIROGENETICS CORP.
XX
XX Paoletti E, Pinc, Pinc, Pincus SE;
XX
XX WPI; 1992-096889/12.
XX
XX Recombinant pox-virus e.g. vaccinia, fowl-pox and canary-pox virus -
XX contg. DNA from flavi-virus e.g. Japanese encephalitis and yellow fever
XX virus, used as vaccine.
XX
XX Example 9; Fig 17; 117pp; English.
XX
XX cDNA was prepared from genomic virion RNA obtained from suspension
XX cultures of C6/36 cells infected with a passage 55 suckling mouse brain
XX stock of the Nagayama strain of JEV. EcoRI linkers were ligated to the
XX cDNA fragments for cloning into pBR322. Recombinant plasmids were
XX transformed into E.coli DH5 cells. Plasmid pC20 contained 81 non-coding
XX nucleotides and the coding regions for C and prM. Sequence AAQ22767 is
XX that of the C coding region of pC20, combined with an updated sequence of
XX the prM, E, NS1, NS2A and NS2B coding regions of the Nagayama strain. The
XX sequence begins at the C protein Met initiation codon. A subfragment of
XX pC20 was cloned into pUC18 to give pDr20. This plasmid was then used in
XX the construction of novel recombinants JEV24, JEV27, JEV33 and JEV34.
XX These were transfected into VP410 infected cells to generate VP825,
XX VP829, VP857 and VP864, respectively
XX
XX Sequence 4512 BP; 1192 A; 1055 C; 1253 G; 1012 T; 0 U; 0 Other;
XX
XX Query Match 54.3%; Score 22.8; DB 2; Length 4512;
XX Best Local Similarity 79.4%; Pred. No. 30;
XX Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
XX
XX 4 CACGCTTGCGCGAGCCCTTACAAGACTATGAAG 37
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 2085 CACGCTGGGCAAGCCCTTTCAACGACTTTGAAG 2118
XX
XX RESULT 51
XX AAN80297
XX ID AAN80297 standard; RNA; 4992 BP.
XX
XX AAN80297;
XX
XX 25-MAR-2003 (revised)
XX 16-SEP-1990 (first entry)
XX
XX Partial nucleic acid sequence of the Japanese encephalitis virus (JEV)
XX genome encoding C, M, E and NS proteins.
XX
XX Japanese encephalitis virus (JEV); vaccine; flavivirus immunity;
XX C protein; M protein; E protein; NS protein.
XX
XX Japanese encephalitis virus.
XX
XX Key Location/Qualifiers
XX mat_peptide 1..159
XX /*tag= a
XX

```

```
FT /product= "C protein"
FT 160..441
FT /tag= b
FT /product= "pre M protein"
FT 442..666
FT /tag= c
FT /product= "M protein"
FT 667..2166
FT /tag= d
FT /product= "E protein"
FT 2167..3408
FT /tag= e
FT /product= "NS1"
FT 3409..3903
FT /tag= f
FT /product= "NS2a"
FT 3904..4296
FT /tag= g
FT /product= "NS2b"
FT 4297..4992
FT /tag= h
FT /product= "NS3"
XX
XX WO8803032-A.
XX
XX 05-MAY-1988.
XX
XX 21-OCT-1987; 87WO-US002763.
XX
XX 27-OCT-1986; 86US-00923907.
XX
XX (FOUR/) FOURNIER M J.
XX
XX Fournier M, Mason TL, Mcada P, Mason PW;
XX WPI; 1988-133144/19.
XX P-PSDB; AAP80264.
XX
XX Nucleic acid of Japanese encephalitis virus - used for producing
XX polypeptide(s) and in the diagnosis of and immunisation against the
XX virus.
XX
XX Claim 4; Fig 5/1-5/5; 52pp; English.
XX
XX Substantially purified nucleic acid comprising at least a 10 bp sequence
XX of DNA corresponding to the nucleic acid sequence of JEV but not to the
XX nucleic acid sequence of yellow fever virus wherein the 10 bp sequence is
XX included within the sequence shown in AAN80297. Pref. the nucleic acid
XX sequences do not correspond to those found in yellow fever virus, West
XX Nile virus, Murray valley fever virus, dengue virus or St Louis
XX encephalitis virus. The nucleic acid may be obtd. from the virus and
XX inserted into a vector. Specific probes can be derived from this nucleic
XX acid. Also claimed are a pure polypeptide synthesized from the purified
XX nucleic acid and antibodies produced to the polypeptide. The polypeptides
XX and nucleic acid prods. are used in the diagnosis of and immunisation
XX against JEV. Inoculation of animals with synthetic viral protein
XX immunogens or with active vectors encoding such immunogens, elicits
XX protective antiviral antibodies. The vaccination induces immunity to the
XX disease caused by other flaviviruses. (Updated on 25-MAR-2003 to correct
XX PA field.) (Updated on 25-MAR-2003 to correct PI field.)
XX
XX Sequence 4992 BP; 1332 A; 1166 C; 1390 G; 3 T; 1100 U; 1 Other;
XX
XX Query Match 54.3%; Score 22.8; DB 1; Length 4992;
XX Best Local Similarity 64.7%; Pred. No. 31;
XX Matches 22; Conservative 5; Mismatches 7; Indels 0; Gaps 0;
XX
XX 4 CACGCTTGCCGAGCCCTTACAAAGACTATGAAG 37
XX ||||| ||| ||||| : ||||| : |||||
XX 1869 CACGCTTGCCGAGCCCTTACAAAGACTATGAAG 1902
XX
XX RESULT 52
XX
XX Query Match 54.3%; Score 22.8; DB 14; Length 7486;
XX Best Local Similarity 79.4%; Pred. No. 34;
XX Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
XX
XX 4 CACGCTTGCCGAGCCCTTACAAAGACTATGAAG 37
XX ||||| ||| ||||| : ||||| : |||||
XX 2710 CACGCTTGCCGAGCCCTTACAAAGACTATGAAG 2743
XX
XX RESULT 53
XX
XX ABK62169
XX ID ABK62169 standard; cDNA; 584 BP.
XX
XX AC ABK62169;
XX
XX DT 18-JUN-2002 (first entry)
XX
XX DE
XX Rat sequence differentially expressed in response to a hepatotoxin #76.
XX
XX Rat; ss; hepatotoxin; expressed sequence tag; EST; drug screening;
XX differential expression; centrilobular necrosis; steatosis.
XX
XX Rattus norvegicus.
XX
XX WO200210453-A2.
XX
XX 07-FEB-2002.
XX
ADW74114
ID ADW74114 standard; DNA; 7486 BP.
XX
XX AC ADW74114;
XX
XX DT 07-APR-2005 (first entry)
XX
XX pcJEME DNA (Japanese encephalitis virus prM/E gene in pcDNA3 vector).
XX
XX nucleic acid vaccine; gene transfer; vector; ds.
XX
XX Japanese encephalitis virus strain Nakayama.
XX Synthetic.
XX
XX JP2005015355-A.
XX
XX 20-JAN-2005.
XX
XX 24-JUN-2003; 2003JP-00179620.
XX
XX 24-JUN-2003; 2003JP-00179620.
XX
XX (UYKO-) UNIV KOBE.
XX
XX WPI; 2005-105264/12.
XX
XX Increasing amount of antigen produced from DNA vaccine, by administering
XX DNA vaccine using needle-less syringe, which produce increased amount of
XX antigen than produced by administering DNA vaccine using usual syringe
XX with needle.
XX
XX Example; SEQ ID NO 1; 22pp; Japanese.
XX
XX The invention relates to a novel method for increasing the amount of
XX antigen produced from a DNA vaccine. The method comprises administering
XX the DNA vaccine using a needle-less syringe, whereby using such a syringe
XX the amount of antigen produced is increased when compared to the amount
XX of antigen produced by administering DNA vaccine using a syringe with a
XX needle. The DNA vaccine is administered by an intramuscular route. The
XX method of the invention may be useful for increasing the amount of
XX antigen produced from a DNA vaccine. The current sequence is that of the
XX pcJEME DNA of the invention which comprises the Japanese encephalitis
XX virus strain Nakayama prM/E gene in the pcDNA3 vector.
XX
XX Sequence 7486 BP; 1791 A; 1909 C; 1948 G; 1838 T; 0 U; 0 Other;
XX
XX Query Match 54.3%; Score 22.8; DB 14; Length 7486;
XX Best Local Similarity 79.4%; Pred. No. 34;
XX Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
XX
XX 4 CACGCTTGCCGAGCCCTTACAAAGACTATGAAG 37
XX ||||| ||| ||||| : ||||| : |||||
XX 2710 CACGCTTGCCGAGCCCTTACAAAGACTATGAAG 2743
XX
XX RESULT 53
XX
XX ABK62169
XX ID ABK62169 standard; cDNA; 584 BP.
XX
XX AC ABK62169;
XX
XX DT 18-JUN-2002 (first entry)
XX
XX DE
XX Rat sequence differentially expressed in response to a hepatotoxin #76.
XX
XX Rat; ss; hepatotoxin; expressed sequence tag; EST; drug screening;
XX differential expression; centrilobular necrosis; steatosis.
XX
XX Rattus norvegicus.
XX
XX WO200210453-A2.
XX
XX 07-FEB-2002.
XX
```

```
XX 30-JUL-2001; 2001WO-US023872.
PP
XX 31-JUL-2000; 2000US-0222040P.
XX 02-NOV-2000; 2000US-0244880P.
PR 11-MAY-2001; 2001US-0290029P.
PR 15-MAY-2001; 2001US-0290645P.
PR 22-MAY-2001; 2001US-0292336P.
PR 06-JUN-2001; 2001US-0295798P.
PR 13-JUN-2001; 2001US-0297457P.
PR 19-JUN-2001; 2001US-0298884P.
PR 09-JUL-2001; 2001US-0303459P.
XX
XX (GENE-) GENE LOGIC INC.
XX
XX Mendrick D, Porter MW, Johnson KR, Castle AL, Elashoff MR;
PI WPI; 2002-241625/29.
XX
XX Predicting toxic effects of compounds or the progression of these toxic
PT effects by determining the changes in gene expression in tissues or cells
PT exposed to the toxin and comparing these to gene expression in unexposed
PT tissues or cells.
XX
XX Claim 1; SEQ ID NO 76; 239pp; English.
XX
XX The invention relates to methods for predicting toxic effects of
CC compounds or the progression of these toxic effects by determining the
CC global changes in gene expression in tissues or cells exposed to the
CC toxin and comparing these to gene expression in unexposed tissues or
CC cells. Also included are methods of predicting at least one toxic effect
CC of a compound or progression of a toxic effect, preferably the
CC hepatotoxicity of a compound, comprising detecting the level of
CC expression in a tissue or cell sample exposed to the compound of two or
CC more genes listed in the specification, where differential expression of
CC the genes is indicative of at least one toxic effect or progression. The
CC method can also be used to identify an agent which modulates the toxic
CC response and predict cellular pathways that a compound modulates in a
CC cell. The methods utilise a set of at least two probes (on a solid
CC support in kit form), where each of the probes comprises a sequence that
CC specifically hybridises to a gene listed in the specification, a computer
CC system comprising a database containing information identifying the
CC expression level in a tissue or cell sample exposed to a hepatotoxin of a
CC set of genes comprising at least two genes listed in the specification,
CC and a user interface to view the information used to present information
CC identifying the expression level in a tissue or cell of at least one gene
CC listed in the specification. The method is useful for elucidating global
CC changes in gene expression and for identifying toxicity markers in
CC tissues or cell exposed to a known toxin. The genes may be used as
CC toxicity markers in drug screening and toxicity assays. The genes and
CC gene expression information may be used as diagnostic markers for the
CC prediction or identification of the physiological state of tissue or cell
CC sample that has been exposed to a compound or agent. Hepatotoxicity is
CC characterised by centrilobular necrosis and steatosis. The present
CC sequence is an expressed sequence tag (EST) or cDNA derived from a gene
CC which is differentially expressed in response to a hepatotoxic agent
XX
XX Sequence 584 BP; 207 A; 108 C; 91 G; 178 T; 0 U; 0 Other;
SQ
Query Match 52.9%; Score 22.2; DB 6; Length 584;
Best Local Similarity 88.9%; Pred. No. 35;
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 15 GAGCCCTTACAAAGACTATGAAGTAA 41
DB 373 GAGCCCTTACAAAGACTATGAAGTAA 399
|||||
RESULT 54
ADB55155
ID ADB55155 standard; DNA; 584 BP.
XX
AC ADB55155;

XX 04-DEC-2003 (first entry)
XX
XX Toxicity-related gene, SEQ ID 181.
XX
XX Toxic; toxin; gene expression profile; hepatotoxicity; liver;
XX drug screening; toxicity assay; ds.
XX
XX Unidentified.
XX
XX WO2003064624-A2.
XX
XX 07-AUG-2003.
XX
XX 31-JAN-2003; 2003WO-US003194.
XX
XX 31-JAN-2002; 2002US-00060087.
PR 15-MAR-2002; 2002US-0364045P.
PR 15-MAR-2002; 2002US-0364055P.
PR 30-DEC-2002; 2002US-0436643P.
XX
XX (GENE-) GENE LOGIC INC.
XX
XX Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Elashoff M;
XX WPI; 2003-589530/65.
XX
XX Predicting a toxic effect of a compound, useful in identifying toxicity
PT markers in liver tissues or cells for drug screening and toxicity assays,
PT comprises preparing gene expression profile of tissue or cells exposed to
PT the compound.
XX
XX Claim 1; SEQ ID NO 181; 1156pp; English.
XX
XX The present invention relates to a method for predicting a toxic effect
CC of a compound. The method comprises preparing a gene expression profile
CC of a tissue or cell sample exposed to the compound, and comparing the
CC gene expression profile to a database comprising SEQ ID 1-4925, where
CC differential expression of the gene indicates at least one toxic effect.
CC The method is useful for predicting at least one toxic effect of a
CC compound, predicting hepatotoxicity or the progression of a toxic effect
CC of a compound, identifying an agent that modulates the onset or
CC progression of a toxic response, predicting the cellular pathways that a
CC compound modulates in a cell, and identifying an agent that modulates at
CC least one activity of a protein. The method and compositions of the
CC present invention using a database of genes having liver toxin-induced
CC differential expression, are useful in identifying toxicity markers in
CC liver tissues or cells for drug screening and toxicity assays. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 584 BP; 207 A; 108 C; 91 G; 178 T; 0 U; 0 Other;
SQ
Query Match 52.9%; Score 22.2; DB 10; Length 584;
Best Local Similarity 88.9%; Pred. No. 35;
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 15 GAGCCCTTACAAAGACTATGAAGTAA 41
DB 373 GAGCCCTTACAAAGACTATGAAGTAA 399
|||||
RESULT 55
ADB49676
ID ADB49676 standard; DNA; 584 BP.
XX
AC ADB49676;

XX 04-DEC-2003 (first entry)
XX
XX Primary rat hepatocyte toxicity modelling related gene SEQ ID NO:218.
XX
```

KW toxic effect; gene expression profile; hepatotoxicity; diagnostic marker;  
KW toxicity marker; toxicity progression; drug screening;  
KW primary rat hepatocyte toxicity modelling; gene; ds.

OS Rattus norvegicus.

PN WO2003065993-A2.

XX 14-AUG-2003.

XX 04-FEB-2003; 2003WO-US003482.

XX 04-FEB-2002; 2002US-0353171P.

PR 13-MAR-2002; 2002US-0363534P.

PR 08-APR-2002; 2002US-0370248P.

PR 10-APR-2002; 2002US-0371134P.

PR 10-APR-2002; 2002US-0371135P.

PR 10-APR-2002; 2002US-0371150P.

PR 11-APR-2002; 2002US-0371413P.

PR 19-APR-2002; 2002US-0373601P.

PR 19-APR-2002; 2002US-0373602P.

PR 22-APR-2002; 2002US-0374139P.

PR 08-MAY-2002; 2002US-0378370P.

PR 09-MAY-2002; 2002US-0378652P.

PR 09-MAY-2002; 2002US-0378653P.

PR 09-MAY-2002; 2002US-0378665P.

PR 09-JUL-2002; 2002US-0394230P.

PR 04-SEP-2002; 2002US-0394253P.

PR 28-JAN-2003; 2003US-0407688P.

XX (GENE-) GENE LOGIC INC.

PA Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Orr M;

XX Elashoff M;

PI WPI; 2003-731472/69.

XX Determining if a compound induces a toxic effect on a tissue or cell, for  
PT identifying hepatotoxic compounds, comprises comparing a gene expression  
PT profile of a tissue or cell sample to a database of Tox mean and non-Tox  
PT mean values.

XX Claim 44; SEQ ID NO 218; 874pp; English.

XX The present invention describes a method for determining whether a  
CC compound induces a toxic effect on a tissue or cell. The method comprises  
CC preparing a gene expression profile of a tissue or cell sample exposed to  
CC the compound, and comparing the gene expression profile to a database  
CC comprising data or information on the Tox mean and non-Tox mean value.  
CC The method is useful for predicting or identifying at least one toxic  
CC effect, particularly hepatotoxicity, of a test or unknown compound. The  
CC genes listed in the specification are useful as diagnostic or toxicity  
CC markers for the prediction or identification of the physiological state  
CC of tissue or cell sample that has been exposed to a compound, or to  
CC identify or predict the toxic effects of a compound or an agent. These  
CC may also be used as markers for monitoring toxicity progression or for  
CC drug screening. The present sequence represents a primary rat hepatocyte  
CC toxicity modelling related gene sequence from the present invention.

XX Sequence 584 BP; 207 A; 108 C; 91 G; 178 T; 0 U; 0 Other;

Query Match 52.9%; Score 22.2; DB 10; Length 584;  
Best Local Similarity 88.9%; Pred. No. 35;  
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 15 GAGCCCTTACAAGACTATGAGTAAA 41

DB 373 GAGCCCTTACAAGACTATGAGTAAA 399

RESULT 56

ADC85956/c

ID ADC85956 standard; DNA; 1339 BP.

XX AC ADC85956;

XX 01-JAN-2004 (first entry)

XX Human GPCR gene SEQ ID NO:409.

XX ds; gene; human; GPCR;

KW guanosine triphosphate-binding protein coupled receptor; gene therapy.

OS Homo sapiens.

XX EP1270724-A2.

XX 02-JAN-2003.

XX 18-JUN-2002; 2002EP-00013517.

XX 18-JUN-2001; 2001JP-00246789.

XX (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.

PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.

XX Suwa M, Asai K, Akiyama Y, Aburatani H;

XX WPI; 2003-315783/31.

DR P-PSDB; ADC85957.

XX New polynucleotide, useful for preparing a composition for treating a

PT patient in need of increased or suppressed activity or expression of the

PT guanosine triphosphate-binding protein coupled receptor.

XX Claim 1; SEQ ID NO 409; 28pp; English.

XX The invention relates to a novel polynucleotide encoding a guanosine  
CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of  
CC the invention may have a use in gene therapy. The polynucleotide and  
CC polypeptide are useful for preparing a composition for treating a patient  
CC in need of increased or suppressed activity or expression of the  
CC guanosine triphosphate-binding protein coupled receptor. The  
CC polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the  
CC invention.

XX Sequence 1339 BP; 288 A; 359 C; 296 G; 396 T; 0 U; 0 Other;

Query Match 51.9%; Score 21.8; DB 10; Length 1339;  
Best Local Similarity 78.8%; Pred. No. 61;  
Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 CACGGTTGGCCGAGCCCTTACAAAGACTATGAA 36

DB 91 CACCCTCTCTGGAGTCCTTACAAAGACAATGAA 59

RESULT 57

ABX63126/c

ID ABX63126 standard; cDNA; 4826 BP.

XX AC ABX63126;

XX 25-FEB-2003 (first entry)

XX Human cDNA #126 differentially expressed in activated vascular tissue.

XX Human; gene; ss; vascular tissue; cytostatic; atherosclerosis; cardiant;  
KW hypotensive; antidiabetic; gynaecological; vasotropic; cerebroprotective;  
KW gene therapy; vascular disease; cancer; coronary; artery disease;  
KW hypertension; diabetes; pre-eclampsia; restenosis;  
KW ischaemia-reperfusion injury; stroke.

OS Homo sapiens.



```
PN US2002137081-A1.
XX
XX
PD 26-SEP-2002.
XX
XX 08-JAN-2002; 2002US-00044090.
XX
XX 28-JUL-2000; 2000US-0222469P.
PR
PR 08-JAN-2001; 2001US-0260483P.
XX
XX (BAND/) BANDMAN O.
PA
XX Bandman O;
XX
XX WPI; 2003-110597/10.
XX
XX Combination for diagnosing, staging, treating, or monitoring the
PT progression of treatment of a vascular disease, e.g. atherosclerosis,
PT comprises several cDNAs that are differentially expressed in activated
PT vascular tissue.
XX
XX Claim 1; Page; 18pp; English.
PS
XX This invention relates to a combination comprising several cDNAs that are
CC differentially expressed in activated vascular tissue. The invention also
CC discloses a high throughput method for detecting differentially expressed
CC cDNAs in a sample. The cDNAs of the invention may have
CC antiarteriosclerotic; cytostatic; cardiac; hypotensive; antidiabetic;
CC gynaecological; vasotropic and cerebroprotective activities and may be
CC used in gene therapy. The cDNAs of the invention may be used in a high-
CC throughput methods for detecting differential expression of one or more
CC cDNAs in a sample, or screening several molecules or compounds to
CC identify a molecule or compound that specifically binds a cDNA of the
CC invention. A protein encoded by the cDNA may be used to screen several
CC molecules or compounds to identify a ligand that specifically binds to
CC the protein, or to produce or purify an antibody to the protein that can
CC be used to detect a protein in a sample or purify a natural or
CC recombinant protein from a sample. The nucleotides may be useful for
CC diagnosing, staging, treating, or monitoring the progression of treatment
CC of a vascular disease, e.g. atherosclerosis, cancer, coronary artery
CC disease, hypertension, diabetes, pre-eclampsia, ischaemia- reperfusion
CC injury, retinosis, or stroke. The cDNAs can also be used for large-scale
CC genetic or gene expression analysis of several new nucleic acid
CC molecules. Antibodies to the proteins encoded by the cDNAs are useful for
CC diagnosing pre-pathologic disorders, and chronic or acute diseases
CC associated with abnormalities in the expression, amount or distribution
CC of the protein. The present sequence represents a cDNA of the invention
CC that is differentially expressed in activated vascular tissue. Note: The
CC sequence data for this patent did not form part of the specification, but
CC was obtained in electronic format directly from USPTO at
CC http.seqdata.uspto.gov/sequence.html?DocID=20020137081
XX
XX Sequence 4826 BP; 1161 A; 1190 C; 1241 G; 1234 T; 0 U; 0 Other;
SQ
Query Match 51.9%; Score 21.8; DB 8; Length 4826;
Best Local Similarity 70.7%; Pred. No. 82;
Matches 29; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
QY 1 GCCACGCTTGCCGAGCCCTTACAAAGACTATGAAGTAAA 41
DB 4091 GCGAAGTTTCACAGAGTCCTTCAAAGACTCTGAAGAAAA 4051
RESULT 58
ADJ56498/c
ID ADJ56498 standard; cDNA; 4826 BP.
XX
XX ADJ56498;
XX
XX 06-MAY-2004 (first entry)
DT
XX Human cDNA differentially expressed in MYCN activated cells SeqID 304.
DE
XX human; differential expression; transactivator; proto-oncogene;
KW
```

```
KW neuroblastoma; small cell lung cancer; cytostatic; gene therapy; ss;
KW MYCN activated cell.
XX
XX Homo sapiens.
OS
XX US2003119009-A1.
PN
XX 26-JUN-2003.
XX
XX 25-FEB-2002; 2002US-00084817.
PF
XX 23-FEB-2001; 2001US-0270784P.
PR
XX (STUA/) STUART S G.
XX (NUCH/) NUCHTERN J G.
PA (PLON/) PLON S E.
PA (SHOH/) SHOHET J M.
XX
XX Stuart SG, Nuchtern JG, Plon SE, Shohet JM;
PI
XX WPI; 2003-635698/60.
DR
XX
XX PT New genes regulated by MYCN activation, useful in gene therapy,
PT particularly for treating a subject with e.g. neuroblastoma or other
PT cancers, or for diagnosing, staging or monitoring the treatment of the
PT cancer.
XX
XX Claim 1; SEQ ID NO 304; 27pp; English.
PS
XX This invention relates to novel isolated cDNAs that are differentially
CC expressed in MYCN activated cells. Specifically, it refers to
CC polynucleotide sequences that exhibit differential expression patterns in
CC cells activated by the transactivator MYCN, where MYCN is a proto-
CC oncogene that is amplified in neuroblastoma cells and is common in small
CC cell lung cancers. The present invention describes these cDNA molecules
CC as useful for in hybridisation assays to detect expression of nucleic
CC acids (or complementary nucleic acids) in a present in a given sample, as
CC well as for screening assays by identifying molecules or compounds that
CC specifically bind the cDNA as a ligand and modulate function or activity.
CC Accordingly, these compositions exhibit cytostatic activity and can also
CC be used for gene therapy purposes. This polynucleotide sequence is a cDNA
CC that is differentially expressed in MYCN activated cells, given in an
CC exemplification of the invention. NOTE: This sequence does not appear in
CC the printed specification but has been obtained in electronic format from
CC the US Patent Office at
CC ftp.seqdata.uspto.gov/sequence.html?DocID=20030119009.
XX
XX Sequence 4826 BP; 1161 A; 1190 C; 1241 G; 1234 T; 0 U; 0 Other;
SQ
Query Match 51.9%; Score 21.8; DB 10; Length 4826;
Best Local Similarity 70.7%; Pred. No. 82;
Matches 29; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
QY 1 GCCACGCTTGCCGAGCCCTTACAAAGACTATGAAGTAAA 41
DB 4091 GCGAAGTTTCACAGAGTCCTTCAAAGACTCTGAAGAAAA 4051
RESULT 59
ABK42428
ID ABK42428 standard; DNA; 17849 BP.
XX
XX ABK42428;
AC
XX
XX 21-MAY-2002 (first entry)
DT
XX Genomic sequence #327 encoding novel human connective tissue polypeptide.
DE
XX Human; connective tissue related disorder; cancer; gene therapy;
KW cytostatic; gene; ds.
XX
XX Homo sapiens.
OS
XX
```





```
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 06-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251898P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465570/50.
XX
XX Isolated nucleic acid molecule encoding a reproductive system antigen is
XX used in preventing, treating or ameliorating a medical condition.
XX
XX Disclosure; SEQ ID NO 10139; 1297pp + Sequence Listing; English.
XX
XX The present invention provides the protein and coding sequences of a
XX number of human reproductive system related antigens. These can be used
XX in the prevention and treatment of reproductive system disorders,
XX including cancer. The present sequence is a genomic sequence encoding a
XX protein of the invention
XX
XX Sequence 17849 BP; 3418 A; 5243 C; 5427 G; 3761 T; 0 U; 0 Other;
SQ
Query Match 51.9%; Score 21.8; DB 4; Length 17849;
Best Local Similarity 70.7%; Pred. No. 1.1e+02;
Matches 29; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
QY 2 CCCACGCTTGGCGGAGCCCTTACAAAGACTATGAAGTAAAG 42
Db 13871 CCCCACACTCTAAGAGGCTTTATAAAACTTTCACGTAAG 13911
. RESULT 61
```

```
ADB60584
ID ADB60584 standard; DNA; 17849 BP.
XX
AC ADB60584;
XX
DT 04-DEC-2003 (first entry)
XX
DE Connective tissue related genomic DNA #327.
XX
XX cytostatic; neuroprotective; nootropic; antiparkinsonian; cardiovascular;
XX antiarteriosclerotic; immunosuppressive; antirheumatic; antiarthritic;
XX antiinflammatory; antiallergic; antiasthmatic; dermatological;
XX nephrotoxic; virucide; fungicide; antibacterial; antiparasitic;
XX gene therapy; ds; connective tissues disorder; rheumatoid arthritis;
XX systemic lupus erythematosus; scleroderma; Sjogren's syndrome; cancer;
XX cancer metastasis; neoplasia; leukaemia; neurodegenerative disorder;
XX Alzheimer's disease; Parkinson's disease; cardiovascular disease;
XX atherosclerosis; myocarditis; cardiopulmonary bypass complication;
XX autoimmune disease; multiple sclerosis; allergic reaction; asthma;
XX rhinitis; eczema; inflammatory condition; Crohn's disease; nephritis;
XX gastrointestinal disorder; inflammatory bowel disease;
XX organ transplant rejection; immune system disorder; Bruton's disease;
XX X-linked lymphoproliferative syndrome;
XX B-cell lymphoproliferative disorder; HIV; AIDS; infection;
XX chromosome identification; chromosome mapping; ds.
XX connective tissue related polynucleotide; gene; ds.
XX
XX Homo sapiens.
XX
XX US2003054375-A1.
XX
XX 20-MAR-2003.
XX
XX 07-MAR-2002; 2002US-00092154.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX 04-FEB-2000; 2000US-0180628P.
XX 24-FEB-2000; 2000US-0184664P.
XX 02-MAR-2000; 2000US-0186350P.
XX 16-MAR-2000; 2000US-0189874P.
XX 17-MAR-2000; 2000US-0190076P.
XX 18-APR-2000; 2000US-0198123P.
XX 19-MAY-2000; 2000US-0205515P.
XX 07-JUN-2000; 2000US-0209467P.
XX 28-JUN-2000; 2000US-0214866P.
XX 30-JUN-2000; 2000US-0215135P.
XX 07-JUL-2000; 2000US-0216647P.
XX 07-JUL-2000; 2000US-0216880P.
XX 11-JUL-2000; 2000US-0217487P.
XX 11-JUL-2000; 2000US-0217496P.
XX 14-JUL-2000; 2000US-0218290P.
XX 26-JUL-2000; 2000US-0220963P.
XX 26-JUL-2000; 2000US-0220964P.
XX 14-AUG-2000; 2000US-0224518P.
XX 14-AUG-2000; 2000US-0224519P.
XX 14-AUG-2000; 2000US-0225213P.
XX 14-AUG-2000; 2000US-0225214P.
XX 14-AUG-2000; 2000US-0225266P.
XX 14-AUG-2000; 2000US-0225267P.
XX 14-AUG-2000; 2000US-0225268P.
XX 14-AUG-2000; 2000US-0225270P.
XX 14-AUG-2000; 2000US-0225447P.
XX 14-AUG-2000; 2000US-0225757P.
XX 14-AUG-2000; 2000US-0225758P.
XX 18-AUG-2000; 2000US-0225759P.
XX 22-AUG-2000; 2000US-0226279P.
XX 22-AUG-2000; 2000US-0226681P.
XX 22-AUG-2000; 2000US-0226686P.
XX 23-AUG-2000; 2000US-0227009P.
XX 30-AUG-2000; 2000US-0228924P.
XX 01-SEP-2000; 2000US-0229287P.
XX 01-SEP-2000; 2000US-0229343P.
```

```
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 02-SEP-2000; 2000US-0229509P.
PR 03-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 13-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 29-SEP-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0246179P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
PR 17-JAN-2001; 2001US-00764847.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM, Barash SC;
XX
XX WPI; 2003-634869/60.
XX P-PSDB; ADS59454.
XX
XX New connective tissue-related polypeptides and polynucleotides, useful
XX for treating, preventing and/or prognosing e.g. disorders of connective
XX tissue, (e.g. rheumatoid arthritis), cancers, cancer metastases and/or
XX neoplasias.
XX
XX Disclosure; SEQ ID NO 1315; 248pp; English.
XX
XX The invention describes an isolated nucleic acid molecule (I), which
XX comprises a sequence that is at least 95 % identical to a connective
XX tissue-related polynucleotide encoding connective tissue antigens (CTA).
XX The polypeptide or polynucleotide is useful for preventing, treating, or
XX ameliorating medical conditions in a mammal. The connective tissue
XX polypeptides, polynucleotides and antibodies are particularly useful for
XX treating, preventing and/or prognosing disorders of connective tissues
XX (e.g. rheumatoid arthritis, discoid and systemic lupus erythematosus,
XX scleroderma, or Sjogren's syndrome), cancers, cancer metastases and/or
XX neoplasias (e.g. leukaemia), neurodegenerative disorders (e.g.
XX Alzheimer's disease, or Parkinson's disease), cardiovascular diseases
XX (e.g. atherosclerosis, myocarditis or cardiopulmonary bypass
XX complications), autoimmune diseases (e.g. systemic lupus erythematosus,
XX rheumatoid arthritis, or multiple sclerosis), allergic reactions (e.g.
XX
Query Match 51.9%; Score 21.8; DB 9; Length 17849;
Best Local Similarity 70.7%; Pred. No. 1.1e+02;
Matches 29; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
QY 2 CCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 42
Db 13871 CCCCACCTTCTAAGAGGCTTATATAAACTTCAAGTAAAG 13911
RESULT-62
ABK42426
ID ABK42426 standard; DNA; 17862 BP.
XX
XX AC ABK42426;
XX
XX DT 21-MAY-2002 (first entry)
XX
```

DE Genomic sequence #325 encoding novel human connective tissue polypeptide.  
XX  
KW Human; connective tissue related disorder; cancer; gene therapy;  
KW cytostatic; gene; ds.  
OS Homo sapiens.  
XX WO20015343-A1.  
PN XX  
PD 02-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US001322.  
XX  
PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234221P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-023935P.  
PR 13-OCT-2000; 2000US-023937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249246P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-024927P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.

```
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX (HUMA-) HUMAN GENOME SCI INC.
PA Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-565190/63.
XX Nucleic acid encoding novel connective tissue associated polypeptides,
PT used in diagnosing, preventing, treating or ameliorating a disorder such
PT as cancer or rheumatoid arthritis.
XX Disclosure; SEQ ID NO 1313; 673pp; English.
XX The present invention relates to the isolation of novel human connective
CC tissue related polypeptides (AAU86435-AAU86923) and the polynucleotide
CC (cDNA and genomic) sequences encoding them. The sequences of the
CC invention are useful in the diagnosis, treatment, prevention and/or
CC prognosis of diseases associated with connective tissue(s), including
CC cancer. The polynucleotide sequences of the invention are also useful in
CC gene therapy. ABK42102-ABK43116 represent genomic sequences encoding the
CC novel human connective tissue related polypeptides. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 17862 BP; 3420 A; 5246 C; 5430 G; 3766 T; 0 U; 0 Other;
Query Match 51.9%; Score 21.8; DB 4; Length 17862;
Best Local Similarity 70.7%; Pred. No. 1.1e+02;
Matches 29; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
QY 2 CCCACGCTTGCGCGAGCCCTTACAAAGACTATGAAGTAAAG 42
Db 13885 CCCCACCTTCTAAGAGGCTTATATAAACTTTCAGTAAAG 13925
RESULT 63
AAL07450
ID AAL07450 standard; DNA; 17862 BP.
XX
AC AAL07450;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen DNA SEQ ID NO: 10138.
XX
KW Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN WO200155320-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US0001339.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235836P.
PR 27-SEP-2000; 2000US-0235837P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
```





PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226686P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 14-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234957P.  
PR 25-SEP-2000; 2000US-0234988P.  
PR 25-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246529P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.

PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 01-DEC-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 06-DEC-2000; 2000US-0256719P.  
PR 08-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 11-DEC-2000; 2000US-0251990P.  
PR 05-JAN-2001; 2001US-0254097P.  
PR 17-JAN-2001; 2001US-00764847.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM, Barash SC;

XX WPI; 2003-634869/60.

DR P-PSDB; ADB59454.

XX New connective tissue-related polypeptides and polynucleotides, useful  
PT for treating, preventing and/or prognosing e.g. disorders of connective  
PT tissue, (e.g. rheumatoid arthritis), cancers, cancer metastases and/or  
PT neoplasias.

XX Disclosure; SEQ ID NO 1313; 248pp; English.

XX The invention describes an isolated nucleic acid molecule (I), which  
XX comprises a sequence that is at least 95 % identical to a connective  
XX tissue-related polynucleotide encoding connective tissue antigens (CTA).  
XX The polypeptide or polynucleotide is useful for preventing, treating, or  
XX ameliorating medical conditions in a mammal. The connective tissue  
XX polypeptides, polynucleotides and antibodies are particularly useful for  
XX treating, preventing and/or prognosing disorders of connective tissues  
XX (e.g. rheumatoid arthritis, discoid and systemic lupus erythematosus,  
XX scleroderma, or Sjogren's syndrome), cancers, cancer metastases and/or  
XX neoplasias (e.g. leukaemia), neurodegenerative disorders (e.g.  
XX Alzheimer's disease, or Parkinson's disease), cardiovascular diseases  
XX (e.g. atherosclerosis, myocarditis or cardiopulmonary bypass  
XX complications), autoimmune diseases (e.g. systemic lupus erythematosus,  
XX rheumatoid arthritis, or multiple sclerosis), allergic reactions (e.g.

Query Match 51.9%; Score 21.8; DB 9; Length 17862;  
Best Local Similarity 70.7%; Pred. No. 1.1e+02;  
Matches 29; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Oy 2 CCCACGCTGGCCGAGCCCTTACAAAGACTATCAAGTAAAG 42

Db 13885 CCCCACCTTCAAGAGGCTTTATAAAACCTTCAAGTAAAG 13925

RESULT 65

```
ABL17076
ID ABL17076 standard; DNA; 21748 BP.
XX
AC ABL17076;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 2701.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
PS Claim 1; SEQ ID NO 2701; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL6175) and the encoded proteins (ABBS57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 21748 BP; 6384 A; 4237 C; 4321 G; 6806 T; 0 U; 0 Other;

Query Match 51.9%; Score 21.8; DB 4; Length 21748;
Best Local Similarity 78.8%; Pred. No. 1.1e+02;
Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 9 TTGGCGGAGCCCTTACAAAGACTATGAAGTAAA 41
|||||
Db 8163 TTGGGTAGCCGTTCAATGACTTTGAATTAAA 8195

RESULT 66
AED85840/c
ID AED85840 standard; DNA; 363 BP.
XX
AC AED85840;
XX
DT 12-JAN-2006 (first entry)
XX
DE Human CD38 antibody heavy chain variable region 3077_VH1B DNA.
XX
KW immunosuppressive; hematological; cytostatic; antiinflammatory;
KW antirheumatic; antiarthritic; antibody; CD38; hematological disease;
KW inflammation; heavy chain variable region; ds; coding sequence.
XX
OS Homo sapiens.

ABL17076
Key Location/Qualifiers
CDS 1..363
FT /*tag= a
FT /partial
FT /product= "3077 VH1B"
FT /note= "No start or stop codon given"
XX
PN WO2005103083-A2.
XX
PD 03-NOV-2005.
XX
PF 07-FEB-2005; 2005WO-IB002476.
XX
PR 06-FEB-2004; 2004US-0541911P.
PR 26-FEB-2004; 2004US-0547584P.
PR 18-MAR-2004; 2004US-0553948P.
PR 06-AUG-2004; 2004US-0599014P.
XX
PA (MORP-) MORPHOSYS AG.
XX
PI Tesar M, Jager U;
XX
DR WPI; 2005-734713/75.
DR P-PSDB; AED85844.
XX
XX New isolated human or humanized antibody or its functional fragment
PT comprising an antigen-binding region that is specific for an epitope of
PT CD38, useful for treating hematological or inflammatory disorders.
XX
PS Claim 62; SEQ ID NO 1; 98pp; English.
XX
CC The invention relates to an isolated human or humanized antibody or its
CC functional fragment comprising an antigen-binding region that is specific
CC for an epitope of CD38. The antibody or its functional fragment is useful
CC for treating a disorder or condition associated with the undesired
CC presence of CD38+ cells, e.g. hematological disease, such as multiple
CC myeloma, chronic lymphocytic leukemia, chronic myelogenous leukemia,
CC acute myelogenous leukemia, and acute lymphocytic leukemia; or an
CC inflammatory disease such as rheumatoid arthritis and systemic lupus
CC erythematosus. The present sequence represents the human CD38 antibody
CC heavy chain variable region 3077_VH1B DNA.
XX
SQ Sequence 363 BP; 85 A; 81 C; 107 G; 90 T; 0 U; 0 Other;

Query Match 51.4%; Score 21.6; DB 14; Length 363;
Best Local Similarity 75.0%; Pred. No. 56;
Matches 27; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 7 GCTTGGCGGAGCCCTTACAAAGACTATGAAGTAAAAG 42
|||||
Db 119 GCTTGGCGGAGCCCAATTATAGATAAGAAAGTAAAAG 84

RESULT 67
AED85870/c
ID AED85870 standard; DNA; 363 BP.
XX
AC AED85870;
XX
DT 12-JAN-2006 (first entry)
XX
DE Human CD38 antibody heavy chain variable region 3077_VH1B DNA.
XX
KW immunosuppressive; hematological; cytostatic; antiinflammatory;
KW antirheumatic; antiarthritic; antibody; CD38; hematological disease;
KW inflammation; heavy chain variable region; ds; coding sequence.
XX
OS Homo sapiens.
XX
PN WO2005103083-A2.
XX
PD 03-NOV-2005.
```

```
PF 07-FEB-2005; 2005WO-IB002476.
XX
XX 06-FEB-2004; 2004US-0541911P.
PR 26-FEB-2004; 2004US-0547584P.
PR 18-MAR-2004; 2004US-0553948P.
PR 06-AUG-2004; 2004US-0559014P.
XX
XX (MORP-) MORPHOSYS AG.
XX
XX Tesar M, Jager U;
XX
XX WPI; 2005-734713/75.
XX
XX New isolated human or humanized antibody or its functional fragment
PT comprising an antigen-binding region that is specific for an epitope of
PT CD38, useful for treating hematological or inflammatory disorders.
XX
XX Disclosure; SEQ ID NO 31; 98pp; English.
XX
XX The invention relates to an isolated human or humanized antibody or its
CC functional fragment comprising an antigen-binding region that is specific
CC for an epitope of CD38. The antibody or its functional fragment is useful
CC for treating a disorder or condition associated with the undesired
CC presence of CD38+ cells, e.g. hematological disease, such as multiple
CC myeloma, chronic lymphocytic leukemia, chronic myelogenous leukemia,
CC acute myelogenous leukemia, and acute lymphocytic leukemia; or an
CC inflammatory disease such as rheumatoid arthritis and systemic lupus
CC erythematosus. The present sequence represents the human CD38 antibody
CC heavy chain variable region 3077_1_VH1B DNA.
XX
XX Sequence 363 BP; 87 A; 80 C; 105 G; 91 T; 0 U; 0 Other;
SQ
Query Match 51.4%; Score 21.6; DB 14; Length 363;
Best Local Similarity 75.0%; Pred. No. 56;
Matches 27; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 7 GCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42
DB 119 GCTTGGCGGACCCCAATTATAGATTAAGAGTAAAG 84
RESULT 68
AAC16864/c
ID AAC16864 standard; cDNA; 452 BP.
XX
XX AAC16864;
XX
XX 06-OCT-2000 (first entry)
XX
XX Human secreted protein 5' EST, SEQ ID NO: 20939.
XX
XX Human, 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX gene therapy; chromosome mapping; ss.
XX
XX Homo sapiens.
XX
XX EP1033401-A2.
XX
XX 06-SEP-2000.
XX
XX 21-FEB-2000; 2000EP-00200610.
XX
XX 26-FEB-1999; 99US-0122487P.
XX
XX (GEST ) GENSET.
XX
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX
XX WPI; 2000-500381/45.
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures.
PT
```

```
XX
PS Claim 1; SEQ ID NO 20939; 71pp + Sequence Listing; English.
XX
XX The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'
CC ESTs are derived from mRNAs with intact 5' ends and can therefore be used
CC to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in
CC diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors
XX
XX Sequence 452 BP; 127 A; 129 C; 109 G; 84 T; 0 U; 3 Other;
SQ
Query Match 51.4%; Score 21.6; DB 3; Length 452;
Best Local Similarity 75.0%; Pred. No. 59;
Matches 27; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 1 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAA 36
DB 400 GTCCAGGCTTGTCAGCCCTTAGAGGGCTAGCAA 365
RESULT 69
AAD42934/c
ID AAD42934 standard; DNA; 70000 BP.
XX
XX AAD42934;
XX
XX 15-NOV-2002 (first entry)
XX
XX Human phospholipase A2 (PLA2), group VI (Ca2+-independent) gene.
XX
XX Human; antisense; phospholipase A2; infection; inflammation; tumour;
XX antisense therapy; PLA2; ds.
XX
XX Homo sapiens.
XX
XX US6410325-B1.
XX
XX 25-JUN-2002.
XX
XX 09-MAY-2001; 2001US-00851896.
XX
XX 09-MAY-2001; 2001US-00851896.
XX
XX (ISIS-) ISIS PHARM INC.
XX
XX Bennett CF, Freier SM, Watt AT;
XX
XX WPI; 2002-616513/66.
XX
XX Novel antisense compounds useful for inhibiting gene expression of human
PT phospholipase A2, group VI and for treating diseases associated with
PT expression of phospholipase A2, group VI.
XX
XX Example 13; Col 49-108; 72pp; English.
XX
XX The present invention relates to novel antisense compounds which inhibit
CC the expression of phospholipase A2 (PLA2), group VI (Ca2+-independent).
CC The invention is useful for inhibiting the expression of PLA2, group VI
CC (Ca2+-independent) in human cells or tissues and for treating an animal,
CC particularly a human suspected of having or being prone to a disease or
CC condition associated with expression of human PLA2, group VI (Ca2+-
CC independent). It is useful for diagnostics, therapeutics and as research
CC reagent, e.g. prophylactically to prevent or delay infection, tumour
CC formation or inflammation. The present DNA sequence is human PLA2, group
```

```
CC VI (Ca2+-independent) gene
XX
SQ Sequence 70000 BP; 15386 A; 18193 C; 17857 G; 18563 T; 0 U; 1 Other;
  Query Match 51.4%; Score 21.6; DB 6; Length 70000;
  Best Local Similarity 75.0%; Pred. No. 1.8e+02;
  Matches 27; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 GCCCAGCTTGCCGAGCCCTTACAAAGACTATGAA 36
  ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 42352 GTCGAGCTTGTCAGCCCTTAGAAGGCTAGCAA 42317

RESULT 70
AAC28686
ID AAC28686 standard; cDNA; 407 BP.
XX
AC AAC28686;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 32761.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW Gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-00200610.
XX
PR 26-FEB-1999; 99US-0122487P.
XX
PA (GEST ) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500381/45.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX
PS Claim 1; SEQ ID NO 32761; 71pp + Sequence Listing; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'
CC ESTs are derived from mRNAs with intact 5' ends and can therefore be used
CC to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in
CC diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors
XX
SQ Sequence 407 BP; 124 A; 78 C; 81 G; 123 T; 0 U; 1 Other;
  Query Match 51.0%; Score 21.4; DB 3; Length 407;
  Best Local Similarity 80.6%; Pred. No. 70;
  Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 11 GCCCGAGCCCTTACAAAGACTATGAAGTAAA 41
  ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 10 GGCATGCGCTTCACAAATGACTATGAAGTAAA 40

CC
XX
SQ Sequence 2356 BP; 650 A; 606 C; 701 G; 399 T; 0 U; 0 Other;
  Query Match 51.0%; Score 21.4; DB 4; Length 2356;
  Best Local Similarity 71.8%; Pred. No. 1e+02;
  Matches 28; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1 GCCCAGCTTGCCGAGCCCTTACAAAGACTATGAAGTA 39
  ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1212 GCCCAAGAATCGCAGAGCCATTATCAAGAAATATCAAGGA 1250

RESULT 72
ADW52673
ID ADW52673 standard; DNA; 882 BP.
XX
AC ADW52673;
XX
DT 21-APR-2005 (first entry)
XX
DE Japanese encephalitis virus outer coat protein coding seq - SEQ ID 2.
XX
KW vaccine; Japanese encephalitis virus infection; outer coat protein; gene;
KW ds.
```



```
PN US2003233675-A1.
XX
XX 18-DEC-2003.
XX
PF 20-FEB-2003; 2003US-00369493.
XX
XX 21-FEB-2002; 2002US-0360039P.
XX
XX (CAOY/) CAO Y.
XX (HINK/) HINKLE G J.
XX (SLAT/) SLATER S C.
XX (CHEN/) CHEN X.
XX (GOLD/) GOLDMAN B S.
XX
XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX WPI; 2004-061375/06.
XX
XX New recombinant DNA construct comprising a promoter positioned to provide
XX for expression of a polynucleotide encoding a polypeptide from a
XX microbial source, useful for producing plants with improved properties.
XX
XX Claim 1; SEQ ID NO 36568; 122pp; English.
XX
XX The invention relates to a recombinant DNA construct comprising a
XX promoter functional in a plant cell, where the promoter is positioned to
XX provide for expression of a polynucleotide encoding a polypeptide from a
XX microbial source. The invention also relates to a transformed plant
XX comprising the recombinant DNA construct and a method of producing a
XX transformed plant having an improved property. The plant is a crop plant
XX such as maize or soybean. The method of producing a transformed plant
XX having an improved property comprises transforming a plant with the
XX recombinant DNA construct and growing the transformed plant, where the
XX polynucleotide or polypeptide is useful for improving plant properties.
XX The recombinant DNA construct is useful for producing plants with
XX improved plant properties, e.g. improved cold, heat or drought tolerance,
XX tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
XX increased resistance to plant disease, better growth rate by modification
XX of the cell cycle pathway with plant growth regulators, increased rate of
XX homologous recombination, modified seed oil or protein yield and/or
XX content, improved yield by modification of carbohydrate, nitrogen or by
XX phosphorus use and/or uptake, by modification of photosynthesis or by
XX providing improved plant growth and development under at least one stress
XX condition, improved lignin production or improved galactomannan
XX production. This sequence represents a bacterial polynucleotide used in
XX the scope of the invention. Note: The sequence data for this patent did
XX not form part of the printed specification but was obtained in electronic
XX format from USPTO at seqdata.uspto.gov/sequence.html.
XX
XX Sequence 1555 BP; 426 A; 371 C; 422 G; 336 T; 0 U; 0 Other;
XX
XX Query Match 50.5%; Score 21.2; DB 13; Length 1555;
XX Best Local Similarity 76.5%; Pred. No. 1.1e+02;
XX Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
XX
QY 2 CCCACGCTTGGCGGAGCCCTTACAAAGACTATGA 35
Db ||| ||| ||| ||| ||| ||| ||| ||| |||
299 CCAACATTGGCGGAGCCCTTACAAAGACTATGA 266
XX
RESULT 75
AD170879
ID AD170879 standard; DNA; 527 BP.
XX
XX AD170879;
XX
XX 20-MAY-2004 (first entry)
XX
XX Human ovarian cancer DNA marker #3621.
XX
XX Human; ovarian cancer; ds; tumour; cytostatic; DNA marker.
XX
XX Homo sapiens.
XX
OS
```

```
XX WO200170979-A2.
XX
XX 27-SEP-2001.
XX
XX 21-MAR-2001; 2001WO-US009126.
XX
XX 21-MAR-2000; 2000US-0191031P.
XX 25-MAY-2000; 2000US-0207124P.
XX 15-JUN-2000; 2000US-0211940P.
XX 07-JUL-2000; 2000US-0216820P.
XX 25-JUL-2000; 2000US-0220661P.
XX 21-DEC-2000; 2000US-0257672P.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX Lee J, Lillie J;
XX
XX WPI; 2001-611502/70.
XX
XX Novel isolated nucleic acid molecules (markers) overexpressed in ovarian
XX cancer cells as compared to their normal non-cancerous ovarian cells are
XX used to characterize stage, grade, histological type of ovarian cancer.
XX
XX Disclosure; SEQ ID NO 3621; 106pp; English.
XX
XX The invention relates to nucleic acid markers which are overexpressed in
XX ovarian cancer cells as compared to their expression in normal (i.e. non-
XX cancerous) ovarian cells. The invention also relates to polypeptides
XX encoded by the markers, antibodies that selectively bind to the
XX polypeptides, a method of inhibiting ovarian cancer in a patient at risk
XX of developing ovarian cancer involving inhibiting expression of a gene
XX corresponding to a marker of the invention and a method of treating a
XX patient afflicted with ovarian cancer comprising providing to cells of
XX the patient an antisense oligonucleotide complementary to a marker of the
XX invention. The markers are useful for assessing if a patient is afflicted
XX with ovarian cancer, which involves comparing the level of expression of
XX a marker in a patient sample and a normal level of expression of the
XX marker in a control non-ovarian cancer sample. A difference between the
XX expression levels indicates ovarian cancer. The level of expression of a
XX marker corresponds to a secreted protein or to a transcribed
XX polynucleotide or its portion. The level of expression of the marker is
XX assessed by detecting the presence in the sample, a protein or protein
XX fragment corresponding to the marker. The presence of protein or protein
XX fragment is detected using an antibody that specifically binds with the
XX protein or protein fragment. Alternatively, the level of expression of
XX the marker is assessed by detecting the presence of a transcribed
XX polynucleotide which anneals with the marker or anneals with a portion of
XX the polynucleotide comprising the marker under stringent conditions. The
XX marker is also used for monitoring the progression of ovarian cancer in a
XX patient which involves detecting expression of the marker in a patient
XX sample at a first point in time, repeating the method at a subsequent
XX time and comparing the level of expression. The method is carried out
XX using an ovarian tissue sample. A composition comprising a marker,
XX polypeptide or antibody of the invention is used to treat ovarian cancer.
XX This sequence represents a human ovarian cancer DNA marker of the
XX invention. Note: The sequence data for this patent did not form part of
XX the printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 527 BP; 187 A; 102 C; 116 G; 120 T; 0 U; 2 Other;
XX
XX Query Match 50.0%; Score 21; DB 5; Length 527;
XX Best Local Similarity 73.0%; Pred. No. 1.1e+02;
XX Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
XX
QY 5 ACGCTTGGCGGAGCCCTTACAAAGACTATGAAGTAAA 41
Db ||| ||| ||| ||| ||| ||| ||| ||| |||
458 AAGTTTCACAGAGTCCTTCAAAAGACTTTGAAGAAAA 494
XX
RESULT 76
AD177193
```

ID ADI77193 standard; DNA; 527 BP.  
XX AC ADI77193;  
XX DT 20-MAY-2004 (first entry)  
XX DE Human ovarian cancer DNA marker #9935.  
XX KW Human; ovarian cancer; ds; tumour; cytostatic; DNA marker.  
XX OS Homo sapiens.  
XX PN WO200170979-A2.  
XX PD 27-SEP-2001.  
XX PF 21-MAR-2001; 2001WO-US009126.  
XX PR 21-MAR-2000; 2000US-0191031P.  
XX PR 25-MAY-2000; 2000US-0207124P.  
XX PR 15-JUN-2000; 2000US-0211940P.  
XX PR 07-JUL-2000; 2000US-0216820P.  
XX PR 25-JUL-2000; 2000US-0220661P.  
XX PR 21-DEC-2000; 2000US-0257672P.  
XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX PI Lee J, Lillie J;  
XX DR WPI; 2001-611502/70.  
XX PT Novel isolated nucleic acid molecules (markers) overexpressed in ovarian  
XX PT cancer cells as compared to their normal non-cancerous ovarian cells are  
XX PT used to characterize stage, grade, histological type of ovarian cancer.  
XX PS Disclosure; SEQ ID NO 9935; 106pp; English.  
XX CC The invention relates to nucleic acid markers which are overexpressed in  
XX CC ovarian cancer cells as compared to their expression in normal (i.e. non-  
XX CC cancerous) ovarian cells. The invention also relates to polypeptides  
XX CC encoded by the markers, antibodies that selectively bind to the  
XX CC polypeptides, a method of inhibiting ovarian cancer in a patient at risk  
XX CC of developing ovarian cancer involving inhibiting expression of a gene  
XX CC corresponding to a marker of the invention and a method of treating a  
XX CC patient afflicted with ovarian cancer comprising providing to cells of  
XX CC the patient an antisense oligonucleotide complementary to a marker of the  
XX CC invention. The markers are useful for assessing if a patient is afflicted  
XX CC with ovarian cancer, which involves comparing the level of expression of  
XX CC a marker in a patient sample and a normal level of expression of the  
XX CC marker in a control non-ovarian cancer sample. A difference between the  
XX CC expression levels indicates ovarian cancer. The level of expression of a  
XX CC marker corresponds to a secreted protein or to a transcribed  
XX CC polynucleotide or its portion. The level of expression of the marker is  
XX CC assessed by detecting the presence in the sample, a protein or protein  
XX CC fragment corresponding to the marker. The presence of protein or protein  
XX CC fragment is detected using an antibody that specifically binds with the  
XX CC protein or protein fragment. Alternatively, the level of expression of  
XX CC the marker is assessed by detecting the presence of a transcribed  
XX CC polynucleotide which anneals with the marker or anneals with a portion of  
XX CC the polynucleotide comprising the marker, under stringent conditions. The  
XX CC marker is also used for monitoring the progression of ovarian cancer in a  
XX CC patient which involves detecting expression of the marker in a patient  
XX CC sample at a first point in time, repeating the method at a subsequent  
XX CC time and comparing the level of expression. The method is carried out  
XX CC using an ovarian tissue sample. A composition comprising a marker,  
XX CC polypeptide or antibody of the invention is used to treat ovarian cancer.  
XX CC This sequence represents a human ovarian cancer DNA marker of the  
XX CC invention. Note: the sequence data for this patent did not form part of  
XX CC the printed specification, but was obtained in electronic format directly  
XX CC from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences).  
XX CC Sequence 527 BP; 187 A; 102 C; 116 G; 120 T; 0 U; 2 Other;

Query Match 50.0%; Score 21; DB 5; Length 527;  
Best Local Similarity 73.0%; Pred. No. 1.1e+02;  
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
QY 5 AGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAA 41  
Db 458 AAGTTTCACAGAGTCCTTCAAAAGACTTTGAAGAAA 494  
RESULT 77  
ADI73027  
ID ADI73027 standard; DNA; 559 BP.  
XX AC ADI73027;  
XX DT 20-MAY-2004 (first entry)  
XX DE Human ovarian cancer DNA marker #5769.  
XX KW Human; ovarian cancer; ds; tumour; cytostatic; DNA marker.  
XX OS Homo sapiens.  
XX PN WO200170979-A2.  
XX PD 27-SEP-2001.  
XX PF 21-MAR-2001; 2001WO-US009126.  
XX PR 21-MAR-2000; 2000US-0191031P.  
XX PR 25-MAY-2000; 2000US-0207124P.  
XX PR 15-JUN-2000; 2000US-0211940P.  
XX PR 07-JUL-2000; 2000US-0216820P.  
XX PR 25-JUL-2000; 2000US-0220661P.  
XX PR 21-DEC-2000; 2000US-0257672P.  
XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX PI Lee J, Lillie J;  
XX DR WPI; 2001-611502/70.  
XX PT Novel isolated nucleic acid molecules (markers) overexpressed in ovarian  
XX PT cancer cells as compared to their normal non-cancerous ovarian cells are  
XX PT used to characterize stage, grade, histological type of ovarian cancer.  
XX PS Disclosure; SEQ ID NO 5769; 106pp; English.  
XX CC The invention relates to nucleic acid markers which are overexpressed in  
XX CC ovarian cancer cells as compared to their expression in normal (i.e. non-  
XX CC cancerous) ovarian cells. The invention also relates to polypeptides  
XX CC encoded by the markers, antibodies that selectively bind to the  
XX CC polypeptides, a method of inhibiting ovarian cancer in a patient at risk  
XX CC of developing ovarian cancer involving inhibiting expression of a gene  
XX CC corresponding to a marker of the invention and a method of treating a  
XX CC patient afflicted with ovarian cancer comprising providing to cells of  
XX CC the patient an antisense oligonucleotide complementary to a marker of the  
XX CC invention. The markers are useful for assessing if a patient is afflicted  
XX CC with ovarian cancer, which involves comparing the level of expression of  
XX CC a marker in a patient sample and a normal level of expression of the  
XX CC marker in a control non-ovarian cancer sample. A difference between the  
XX CC expression levels indicates ovarian cancer. The level of expression of a  
XX CC marker corresponds to a secreted protein or to a transcribed  
XX CC polynucleotide or its portion. The level of expression of the marker is  
XX CC assessed by detecting the presence in the sample, a protein or protein  
XX CC fragment corresponding to the marker. The presence of protein or protein  
XX CC fragment is detected using an antibody that specifically binds with the  
XX CC protein or protein fragment. Alternatively, the level of expression of  
XX CC the marker is assessed by detecting the presence of a transcribed  
XX CC polynucleotide which anneals with the marker or anneals with a portion of  
XX CC the polynucleotide comprising the marker, under stringent conditions. The  
XX CC marker is also used for monitoring the progression of ovarian cancer in a  
XX CC patient which involves detecting expression of the marker in a patient  
XX CC sample at a first point in time, repeating the method at a subsequent  
XX CC time and comparing the level of expression. The method is carried out  
XX CC using an ovarian tissue sample. A composition comprising a marker,  
XX CC polypeptide or antibody of the invention is used to treat ovarian cancer.  
XX CC This sequence represents a human ovarian cancer DNA marker of the  
XX CC invention. Note: the sequence data for this patent did not form part of  
XX CC the printed specification, but was obtained in electronic format directly  
XX CC from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences).  
XX CC Sequence 527 BP; 187 A; 102 C; 116 G; 120 T; 0 U; 2 Other;





CC of developing ovarian cancer involving inhibiting expression of a gene  
CC corresponding to a marker of the invention and a method of treating a  
CC patient afflicted with ovarian cancer comprising providing to cells of  
CC the patient an antisense oligonucleotide complementary to a marker of the  
CC invention. The markers are useful for assessing if a patient is afflicted  
CC with ovarian cancer, which involves comparing the level of expression of  
CC a marker in a patient sample and a normal level of expression of the  
CC marker in a control non-ovarian cancer sample. A difference between the  
CC expression levels indicates ovarian cancer. The level of expression of a  
CC marker corresponds to a secreted protein or to a transcribed  
CC polynucleotide or its portion. The level of expression of the marker is  
CC assessed by detecting the presence in the sample, a protein or protein  
CC fragment corresponding to the marker. The presence of protein or protein  
CC fragment is detected using an antibody that specifically binds with the  
CC protein or protein fragment. Alternatively, the level of expression of  
CC the marker is assessed by detecting the presence of a transcribed  
CC polynucleotide which anneals with the marker or anneals with a portion of  
CC the polynucleotide comprising the marker, under stringent conditions. The  
CC marker is also used for monitoring the progression of ovarian cancer in a  
CC patient which involves detecting expression of the marker in a patient  
CC sample at a first point in time, repeating the method at a subsequent  
CC time and comparing the level of expression. The method is carried out  
CC using an ovarian tissue sample. A composition comprising a marker,  
CC polypeptide or antibody of the invention is used to treat ovarian cancer.  
CC This sequence represents a human ovarian cancer DNA marker of the  
CC invention. Note: The sequence data for this patent did not form part of  
CC the printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 579 BP; 205 A; 114 C; 134 G; 125 T; 0 U; 1 Other;

Query Match 50.0%; Score 21; DB 5; Length 579;  
Best Local Similarity 73.0%; Pred. No. 1.1e+02;  
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 5 ACCTTTGGCGGAGCCCTTACAAAGACTATGAAGTAAA 41  
Db 459 AAGTTTCACAGAGTCCTTCAAAAGACTTTGAAGAAAA 495

RESULT 80

AD171451

ID AD171451 standard; DNA; 579 BP.

XX AC AD171451;

XX DT 20-MAY-2004 (first entry)

XX DE Human ovarian cancer DNA marker #4193.

XX KW Human; ovarian cancer; ds; tumour; cytostatic; DNA marker.

XX OS Homo sapiens.

XX PN WO200170979-A2.

XX PD 27-SEP-2001.

XX PF 21-MAR-2001; 2001WO-US0009126.

XX PR 21-MAR-2000; 2000US-0191031P.

XX PR 25-MAY-2000; 2000US-0207124P.

XX PR 15-JUN-2000; 2000US-0211940P.

XX PR 07-JUL-2000; 2000US-0216820P.

XX PR 25-JUL-2000; 2000US-0220661P.

XX PR 21-DEC-2000; 2000US-0257672P.

XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX PI Lee J, Lillie J;

XX DR WPI; 2001-611502/70.

PT Novel isolated nucleic acid molecules (markers) overexpressed in ovarian  
PT cancer cells as compared to their normal non-cancerous ovarian cells are  
PT used to characterize stage, grade, histological type of ovarian cancer.

XX Disclosure; SEQ ID NO 4193; 106pp; English.

XX CC The invention relates to nucleic acid markers which are overexpressed in  
XX CC ovarian cancer cells as compared to their expression in normal (i.e. non-  
XX CC cancerous) ovarian cells. The invention also relates to polypeptides  
XX CC encoded by the markers, antibodies that selectively bind to the  
XX CC polypeptides, a method of inhibiting ovarian cancer in a patient at risk  
XX CC of developing ovarian cancer involving inhibiting expression of a gene  
XX CC corresponding to a marker of the invention and a method of treating a  
XX CC patient afflicted with ovarian cancer comprising providing to cells of  
XX CC the patient an antisense oligonucleotide complementary to a marker of the  
XX CC invention. The markers are useful for assessing if a patient is afflicted  
XX CC with ovarian cancer, which involves comparing the level of expression of  
XX CC a marker in a patient sample and a normal level of expression of the  
XX CC marker in a control non-ovarian cancer sample. A difference between the  
XX CC expression levels indicates ovarian cancer. The level of expression of a  
XX CC marker corresponds to a secreted protein or to a transcribed  
XX CC polynucleotide or its portion. The level of expression of the marker is  
XX CC assessed by detecting the presence in the sample, a protein or protein  
XX CC fragment corresponding to the marker. The presence of protein or protein  
XX CC fragment is detected using an antibody that specifically binds with the  
XX CC protein or protein fragment. Alternatively, the level of expression of  
XX CC the marker is assessed by detecting the presence of a transcribed  
XX CC polynucleotide which anneals with the marker or anneals with a portion of  
XX CC the polynucleotide comprising the marker, under stringent conditions. The  
XX CC marker is also used for monitoring the progression of ovarian cancer in a  
XX CC patient which involves detecting expression of the marker in a patient  
XX CC sample at a first point in time, repeating the method at a subsequent  
XX CC time and comparing the level of expression. The method is carried out  
XX CC using an ovarian tissue sample. A composition comprising a marker,  
XX CC polypeptide or antibody of the invention is used to treat ovarian cancer.  
XX CC This sequence represents a human ovarian cancer DNA marker of the  
XX CC invention. Note: The sequence data for this patent did not form part of  
XX CC the printed specification, but was obtained in electronic format directly  
XX CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 579 BP; 205 A; 114 C; 134 G; 125 T; 0 U; 1 Other;

Query Match 50.0%; Score 21; DB 5; Length 579;

Best Local Similarity 73.0%; Pred. No. 1.1e+02;  
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 5 ACCTTTGGCGGAGCCCTTACAAAGACTATGAAGTAAA 41  
Db 459 AAGTTTCACAGAGTCCTTCAAAAGACTTTGAAGAAAA 495

RESULT 81

ADL44546

ID ADL44546 standard; DNA; 625 BP.

XX AC ADL44546;

XX DT 20-MAY-2004 (first entry)

XX DE Human ovarian cancer DNA marker #18436.

XX KW Human; ovarian cancer; ds; tumour; cytostatic; DNA marker.

XX OS Homo sapiens.

XX PN WO200170979-A2.

XX PD 27-SEP-2001.

XX PF 21-MAR-2001; 2001WO-US0009126.

XX PR 21-MAR-2000; 2000US-0191031P.

XX PR 25-MAY-2000; 2000US-0207124P.

```

PR 15-JUN-2000; 2000US-0211940P.
PR 07-JUL-2000; 2000US-0216820P.
PR 25-JUL-2000; 2000US-0220661P.
PR 21-DEC-2000; 2000US-0257672P.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX Lee J, Lillie J;
XX WPI; 2001-611502/70.
XX
XX Novel isolated nucleic acid molecules (markers) overexpressed in ovarian
XX cancer cells as compared to their normal non-cancerous ovarian cells are
XX used to characterize stage, grade, histological type of ovarian cancer.
XX
XX Disclosure; SEQ ID NO 18436; 106pp; English.
XX
XX The invention relates to nucleic acid markers which are overexpressed in
XX ovarian cancer cells as compared to their expression in normal (i.e. non-
XX cancerous) ovarian cells. The invention also relates to polypeptides
XX encoded by the markers, antibodies that selectively bind to the
XX polypeptides, a method of inhibiting ovarian cancer in a patient at risk
XX of developing ovarian cancer involving inhibiting expression of a gene
XX corresponding to a marker of the invention and a method of treating a
XX patient afflicted with ovarian cancer comprising providing to cells of
XX the patient an antisense oligonucleotide complementary to a marker of the
XX invention. The markers are useful for assessing if a patient is afflicted
XX with ovarian cancer, which involves comparing the level of expression of
XX a marker in a patient sample and a normal level of expression of the
XX expression levels indicates ovarian cancer. The level of expression of a
XX marker corresponds to a secreted protein or to a transcribed
XX polynucleotide or its portion. The level of expression of the marker is
XX assessed by detecting the presence in the sample, a protein or protein
XX fragment corresponding to the marker. The presence of protein or protein
XX fragment is detected using an antibody that specifically binds with the
XX protein or protein fragment. Alternatively, the level of expression of
XX the marker is assessed by detecting the presence of a transcribed
XX polynucleotide which anneals with the marker or anneals with a portion of
XX the polynucleotide comprising the marker, under stringent conditions. The
XX marker is also used for monitoring the progression of ovarian cancer in a
XX patient which involves detecting expression of the marker in a patient
XX sample at a first point in time, repeating the method at a subsequent
XX time and comparing the level of expression. The method is carried out
XX using an ovarian tissue sample. A composition comprising a marker,
XX polypeptide or antibody of the invention is used to treat ovarian cancer.
XX This sequence represents a human ovarian cancer DNA marker of the
XX invention.
XX
XX Sequence 625 BP; 214 A; 128 C; 144 G; 138 T; 0 U; 1 Other;
XX
XX Query Match 50.0%; Score 21; DB 5; Length 625;
XX Best Local Similarity 73.0%; Pred. No. 1.1e+02;
XX Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
XX
QY 5 AGCTTGCCGAGCCCTTCAAGACTATGAAGTAAA 41
DB 538 AAGTTTCACAGAGTCCTTCAAAAGACTCTGAAGAAA 574
XX
RESULT 82
ADL44897/c
ID ADL44897 standard; DNA; 692 BP.
XX
XX AC ADL44897;
XX
XX 20-WAY-2004 (first entry)
XX
XX Human ovarian cancer DNA marker #18787.
XX
XX Human; ovarian cancer; ds; tumour; cytostatic; DNA marker.
XX
XX Homo sapiens.
XX

```

---

```

XX WO200170979-A2.
XX
XX 27-SEP-2001.
XX
XX 21-MAR-2001; 2001WO-US009126.
XX
XX 21-MAR-2000; 2000US-0191031P.
XX 25-MAY-2000; 2000US-0207124P.
XX 15-JUN-2000; 2000US-0211940P.
XX 07-JUL-2000; 2000US-0216820P.
XX 25-JUL-2000; 2000US-0220661P.
XX 21-DEC-2000; 2000US-0257672P.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX Lee J, Lillie J;
XX WPI; 2001-611502/70.
XX
XX Novel isolated nucleic acid molecules (markers) overexpressed in ovarian
XX cancer cells as compared to their normal non-cancerous ovarian cells are
XX used to characterize stage, grade, histological type of ovarian cancer.
XX
XX Disclosure; SEQ ID NO 18787; 106pp; English.
XX
XX The invention relates to nucleic acid markers which are overexpressed in
XX ovarian cancer cells as compared to their expression in normal (i.e. non-
XX cancerous) ovarian cells. The invention also relates to polypeptides
XX encoded by the markers, antibodies that selectively bind to the
XX polypeptides, a method of inhibiting ovarian cancer in a patient at risk
XX of developing ovarian cancer involving inhibiting expression of a gene
XX corresponding to a marker of the invention and a method of treating a
XX patient afflicted with ovarian cancer comprising providing to cells of
XX the patient an antisense oligonucleotide complementary to a marker of the
XX invention. The markers are useful for assessing if a patient is afflicted
XX with ovarian cancer, which involves comparing the level of expression of
XX a marker in a patient sample and a normal level of expression of the
XX expression levels indicates ovarian cancer. The level of expression of a
XX marker corresponds to a secreted protein or to a transcribed
XX polynucleotide or its portion. The level of expression of the marker is
XX assessed by detecting the presence in the sample, a protein or protein
XX fragment corresponding to the marker. The presence of protein or protein
XX fragment is detected using an antibody that specifically binds with the
XX protein or protein fragment. Alternatively, the level of expression of
XX the marker is assessed by detecting the presence of a transcribed
XX polynucleotide which anneals with the marker or anneals with a portion of
XX the polynucleotide comprising the marker, under stringent conditions. The
XX marker is also used for monitoring the progression of ovarian cancer in a
XX patient which involves detecting expression of the marker in a patient
XX sample at a first point in time, repeating the method at a subsequent
XX time and comparing the level of expression. The method is carried out
XX using an ovarian tissue sample. A composition comprising a marker,
XX polypeptide or antibody of the invention is used to treat ovarian cancer.
XX This sequence represents a human ovarian cancer DNA marker of the
XX invention.
XX
XX Sequence 692 BP; 154 A; 173 C; 138 G; 227 T; 0 U; 0 Other;
XX
XX Query Match 50.0%; Score 21; DB 5; Length 692;
XX Best Local Similarity 73.0%; Pred. No. 1.2e+02;
XX Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
XX
QY 5 AGCTTGCCGAGCCCTTCAAGACTATGAAGTAAA 41
DB 521 AAGTTTCACAGAGTCCTTCAAAAGACTCTGAAGAAA 485
XX
RESULT 83
AD030325

```

ID AD030325 standard; cDNA; 1068 BP.  
XX AC AD030325;  
XX DT 29-JUL-2004 (first entry)  
XX DE Mouse GPCR RAI3 polynucleotide, SEQ ID NO:1428.  
XX DX  
XX KW G protein-coupled receptor; GPCR; drug screening; diagnosis;  
KW transgenic mouse; neurological disorder; adrenal gland disorder;  
KW colon disorder; intestinal disorder; cardiovascular disorder;  
KW muscular disorder; blood disorder; immune disorder; bone disorder;  
KW joint disorder; metabolic disorder; nutritive disorder; cancer;  
KW kidney disorder; liver disorder; lung disorder; breast disorder;  
KW ovary disorder; uterus disorder; prostate disorder; testis disorder;  
KW skin disorder; stomach disorder; pancreas disorder; spleen disorder;  
KW thymus disorder; thyroid disorder; antiparkinsonian; antinemic;  
KW cytostatic; antiinflammatory; vasotropic; antianginal; antiarrhythmic;  
KW CNS; central nervous system; respiratory; antidiarrhoeic; antidiabetic;  
KW virucide; hepatotropic; antibacterial; antianaemic; antiseborrhoeic;  
KW dermatological; antiulcer; antithyroid; antiallergic; anorectic;  
KW immunosuppressive; nephrotropic; gene therapy; GPCR modulator; mouse;  
KW murine; gene; ss.  
XX OS Mus musculus.  
XX PN WO2004040000-A2.  
XX PD 13-MAY-2004.  
XX XX 09-SEP-2003; 2003WO-US028226.  
XX PF 09-SEP-2002; 2002US-0409303P.  
XX PR 09-APR-2003; 2003US-0461329P.  
XX XX (PRIM-) PRIMAL INC.  
XX PI Gaitanaris GA, Bergmann JE, Gragerov A, Hohmann J, Li F;  
PI Madisen L, McIlwain KL, Pavlova MN, Vasiliadis D, Zeng H;  
XX WPI; 2004-390329/36.  
XX DR P-PSDB; ADO29633.  
XX XX Novel mammalian G protein coupled receptors, useful for identifying  
PT compounds that modulates diagnosing and treating disease condition  
PT associated with GPCR dysfunction e.g. autoimmune diseases, angina  
PT pectoris, Parkinson's disease.  
XX XX Claim 151; SEQ ID NO 1428; 542pp; English.  
XX PS  
XX CC The invention relates to human and mouse G protein-coupled receptors  
CC (GPCRs) and nucleic acids encoding them. The invention also relates to  
CC sequences at least 90% identical to the GPCR proteins and nucleic acids  
CC of the invention; methods of treating, preventing or diagnosing diseases  
CC associated with GPCRs of the invention; methods of screening for  
CC compounds useful in the treatment of GPCR-related diseases; a transgenic  
CC mouse comprising a GPCR gene of the invention; a mouse comprising a  
CC mutation in a GPCR transgene or in an endogenous GPCR gene; cells derived  
CC from the transgenic mice; kits comprising several mice, each of which has  
CC a mutation in a different GPCR gene of the invention; and kits comprising  
CC probes which hybridise to GPCR polynucleotides of the invention. The  
CC invention further discloses variants of the GPCR polypeptides and vectors  
CC comprising a GPCR nucleic acid. The GPCR nucleic acids and proteins may  
CC be used in the diagnosis, treatment or prevention of a wide variety of  
CC diseases including neurological disorders (e.g., Alzheimer's disease,  
CC depression, diabetic neuropathy, Parkinson's disease or schizophrenia);  
CC disorders of the adrenal gland; disorders of the colon or intestine  
CC (e.g., Crohn's disease, diarrhoea, food poisoning or irritable bowel  
CC syndrome); cardiovascular disorders (e.g., angina, cardiac arrhythmia or  
CC myocardial infarction); muscular disorders; blood disorders (e.g.,  
CC anaemia or leukaemia); immune disorders (e.g., autoimmune disorders or  
CC AIDS); bone and joint disorders (e.g., osteoarthritis, rheumatoid  
CC arthritis, gout or osteoporosis); metabolic or nutritive disorders (e.g.,  
CC obesity, enzyme deficiency-related diseases or vitamin deficiency-related  
CC diseases); and disorders of the kidney, liver, lung, breast, ovary,  
CC uterus, prostate, testis, skin, stomach, pancreas, spleen, thymus and  
CC thyroid (e.g., cancers). The present sequence represents a GPCR-encoding  
CC nucleic acid of the invention. Note: The full sequence data for this  
CC patent did not form part of the printed specification; those sequences  
CC not shown were obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX SQ Sequence 1068 BP; 205 A; 319 C; 267 G; 277 T; 0 U; 0 Other;  
XX Query Match 50.0%; Score 21; DB 12; Length 1068;  
XX Best Local Similarity 73.0%; Pred. No. 1.3e+02;  
XX Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
XX QY 1 GCCACGCTTGCCGAGCCCTTACAAAGACTATGAAG 37  
XX Db 1015 GCCACGCCCCGCCAGTCGCTACATGACTACGAAG 1051  
XX RESULT 84  
XX ADL62421  
XX ID ADL62421 standard; DNA; 1096 BP.  
XX AC ADL62421;  
XX DT 20-MAY-2004 (first entry)  
XX DE Human ovarian cancer DNA marker #20633.  
XX KW Human; ovarian cancer; ds; tumour; cytostatic; DNA marker.  
XX OS Homo sapiens.  
XX PN WO200170979-A2.  
XX PD 27-SEP-2001.  
XX XX 21-MAR-2001; 2001WO-US009126.  
XX PR 21-MAR-2000; 2000US-0191031P.  
XX PR 25-MAY-2000; 2000US-0207124P.  
XX PR 15-JUN-2000; 2000US-0211940P.  
XX PR 07-JUL-2000; 2000US-0216820P.  
XX PR 25-JUL-2000; 2000US-0220661P.  
XX PR 21-DEC-2000; 2000US-0257672P.  
XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX PI Lee J, Lillie J;  
XX WPI; 2001-611502/70.  
XX PT Novel isolated nucleic acid molecules (markers) overexpressed in ovarian  
PT cancer cells as compared to their normal non-cancerous ovarian cells are  
PT used to characterize stage, grade, histological type of ovarian cancer.  
XX PS Disclosure; SEQ ID NO 20633; 106pp; English.  
XX CC The invention relates to nucleic acid markers which are overexpressed in  
CC ovarian cancer cells as compared to their expression in normal (i.e. non-  
CC cancerous) ovarian cells. The invention also relates to polypeptides  
CC encoded by the markers, antibodies that selectively bind to the  
CC polypeptides, a method of inhibiting ovarian cancer in a patient at risk  
CC of developing ovarian cancer involving inhibiting expression of a gene  
CC corresponding to a marker of the invention and a method of treating a  
CC patient afflicted with ovarian cancer comprising providing to cells of  
CC the patient an antisense oligonucleotide complementary to a marker of the  
CC invention. The markers are useful for assessing if a patient is afflicted  
CC with ovarian cancer, which involves comparing the level of expression of  
CC a marker in a patient sample and a normal level of expression of the  
CC marker in a control non-ovarian cancer sample. A difference between the  
CC expression levels indicates ovarian cancer. The level of expression of a

CC marker corresponds to a secreted protein or to a transcribed  
CC polynucleotide or its portion. The level of expression of the marker is  
CC assessed by detecting the presence in the sample, a protein or protein  
CC fragment corresponding to the marker. The presence of protein or protein  
CC fragment is detected using an antibody that specifically binds with the  
CC protein or protein fragment. Alternatively, the level of expression of  
CC the marker is assessed by detecting the presence of a transcribed  
CC polynucleotide which anneals with the marker or anneals with a portion of  
CC the polynucleotide comprising the marker, under stringent conditions. The  
CC marker is also used for monitoring the progression of ovarian cancer in a  
CC patient which involves detecting expression of the marker in a patient  
CC sample at a first point in time, repeating the method at a subsequent  
CC time and comparing the level of expression. The method is carried out  
CC using an ovarian tissue sample. A composition comprising a marker,  
CC polypeptide or antibody of the invention is used to treat ovarian cancer.  
CC This sequence represents a human ovarian cancer DNA marker of the  
CC invention. Note: The sequence data for this patent did not form part of  
CC the printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 1096 BP; 367 A; 215 C; 286 G; 224 T; 0 U; 4 Other;

Query Match 50.0%; Score 21; DB 5; Length 1096;  
Best Local Similarity 73.0%; Pred. No. 1.3e+02;  
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
QY 5 AGCGTTGGCCGAGCCCTTACAAAGACTATGAAGTAAA 41  
DB 437 AAGTTTCACAGAGTCTCTCAAAAGACTCTGAAGAAAA 473  
|||||  
|||||

RESULT 85  
ADL62598  
ID ADL62598 standard; DNA; 1386 BP.  
XX AC ADL62598;  
XX DT 20-MAY-2004 (first entry)  
XX DE Human ovarian cancer DNA marker #20810.  
XX KW Human; ovarian cancer; ds; tumour; cytostatic; DNA marker.  
XX OS Homo sapiens.  
XX PN WO200170979-A2.  
XX PD 27-SEP-2001.  
XX PF 21-MAR-2001; 2001WO-US009126.

XX PR 21-MAR-2000; 2000US-0191031P.  
XX PR 25-MAY-2000; 2000US-0207124P.  
XX PR 15-JUN-2000; 2000US-0211940P.  
XX PR 07-JUL-2000; 2000US-0216820P.  
XX PR 25-JUL-2000; 2000US-0220661P.  
XX PR 21-DEC-2000; 2000US-0257672P.  
XX PA (WILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX PI Lee J, Lillie J;  
XX DR WPI; 2001-611502/70.  
XX PT Novel isolated nucleic acid molecules (markers) overexpressed in ovarian  
XX PT cancer cells as compared to their normal non-cancerous ovarian cells are  
XX PT used to characterize stage, grade, histological type of ovarian cancer.  
XX PS Disclosure; SEQ ID NO 20810; 106pp; English.  
XX CC The invention relates to nucleic acid markers which are overexpressed in  
XX CC ovarian cancer cells as compared to their expression in normal (i.e. non-  
XX CC cancerous) ovarian cells. The invention also relates to polypeptides

CC encoded by the markers, antibodies that selectively bind to the  
CC polypeptides, a method of inhibiting ovarian cancer in a patient at risk  
CC of developing ovarian cancer involving inhibiting expression of a gene  
CC corresponding to a marker of the invention and a method of treating a  
CC patient afflicted with ovarian cancer comprising providing to cells of  
CC the patient an antisense oligonucleotide complementary to a marker of the  
CC invention. The markers are useful for assessing if a patient is afflicted  
CC with ovarian cancer, which involves comparing the level of expression of  
CC a marker in a patient sample and a normal level of expression of the  
CC marker in a control non-ovarian cancer sample. A difference between the  
CC expression levels indicates ovarian cancer. The level of expression of a  
CC marker corresponds to a secreted protein or to a transcribed  
CC polynucleotide or its portion. The level of expression of the marker is  
CC assessed by detecting the presence in the sample, a protein or protein  
CC fragment corresponding to the marker. The presence of protein or protein  
CC fragment is detected using an antibody that specifically binds with the  
CC protein or protein fragment. Alternatively, the level of expression of  
CC the marker is assessed by detecting the presence of a transcribed  
CC polynucleotide which anneals with the marker or anneals with a portion of  
CC the polynucleotide comprising the marker, under stringent conditions. The  
CC marker is also used for monitoring the progression of ovarian cancer in a  
CC patient which involves detecting expression of the marker in a patient  
CC sample at a first point in time, repeating the method at a subsequent  
CC time and comparing the level of expression. The method is carried out  
CC using an ovarian tissue sample. A composition comprising a marker,  
CC polypeptide or antibody of the invention is used to treat ovarian cancer.  
CC This sequence represents a human ovarian cancer DNA marker of the  
CC invention.  
XX  
XX Sequence 1386 BP; 457 A; 287 C; 338 G; 300 T; 0 U; 4 Other;

Query Match 50.0%; Score 21; DB 5; Length 1386;  
Best Local Similarity 73.0%; Pred. No. 1.4e+02;  
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
QY 5 AGCGTTGGCCGAGCCCTTACAAAGACTATGAAGTAAA 41  
DB 566 AAGTTTCACAGAGTCTCTCAAAAGACTCTGAAGAAAA 602  
|||||  
|||||

RESULT 86  
ADS10225  
ID ADS10225 standard; DNA; 1566 BP.  
XX AC ADS10225;  
XX DT 16-DEC-2004 (first entry)  
XX DE Human therapeutic DNA - SEQ ID 462.  
XX KW antinflammatory; neuroprotective; antianaemic; cytostatic; vulnary;  
XX KW inflammatory; haematopoiesis; immunity; neurodegenerative; stem cell;  
XX KW aplastic anaemia; cancer; wound healing; gene therapy; ds; gene.  
XX OS Homo sapiens.  
XX PN WO2004080148-A2.  
XX PD 23-SEP-2004.  
XX PF 30-SEP-2003; 2003WO-US030720.  
XX PR 02-OCT-2002; 2002US-0416186P.  
XX PA (NUVE-) NUVELO INC.  
XX PI Tang YT, Asundi V, Ren F, Zhang J, Zhang J, Wehrman T, Wang Z, Ma Y;  
XX PI Wang D, Chen R, Zhao QA, Wang J, Ghosh M, Xue AJ, Weng G, Zhou P;  
XX WPI; 2004-668857/65.  
XX DR P-PSDB; ADS10909.  
XX PT New polynucleotide, useful in preparing a composition for diagnosing or

CC mammals. The invention also relates to nucleic acid and polypeptide  
CC sequences at least 80% identical to the TAT nucleic acids and  
CC polypeptides; expression vectors and host cells comprising a TAT nucleic  
CC acid; an antibody specific for a TAT polypeptide; a peptide or organic  
CC molecule which binds to a TAT polypeptide; fusion proteins comprising a  
CC TAT polypeptide; and methods and compositions for the treatment or  
CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,  
CC antibodies, antagonists, binding molecules and compositions are useful  
CC for diagnosing or treating a cell proliferative disorder associated with  
CC increased TAT expression, particularly cancers such as breast cancer,  
CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder  
CC cancer, pancreatic cancer, cervical cancer, cancers of the central  
CC nervous system, melanoma and leukaemia. TAT nucleic acids may further be  
CC used as hybridisation probes, in chromosome and gene mapping, in  
CC chromosome identification and in gene therapy. The present sequence  
CC represents a TAT nucleic acid of the invention  
XX  
SQ Sequence 1580 BP; 486 A; 348 C; 404 G; 342 T; 0 U; 0 Other;  
  
Query Match 50.0%; Score 21; DB 13; Length 1580;  
Best Local Similarity 73.0%; Pred. No. 1.4e+02;  
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
  
QY 5 ACCTTGGCGAGCCCTTACAAAGACTATGAAGTAAA 41  
| | | | | | | | | | | | | | | | | | | | |  
DB 557 AAGTTTCACAGAGTCTCTTCAAAAGACTCTGAAGAAAA 593  
  
RESULT 88  
ABN59642  
ID ABN59642 standard; cDNA; 1648 BP.  
XX AC ABN59642;  
XX  
DT 28-JUN-2002 (first entry)  
XX  
DE Novel human coding sequence SEQ ID NO: 53.  
XX  
KW Human; antianaemic; vulnery; antiinflammatory; immunomodulator;  
KW antiinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;  
KW neuroprotective; antiparkinsonian; protein therapy; EST;  
KW expressed sequence tag; gene; ss.  
XX  
OS Homo sapiens.  
XX  
XX WO200222660-A2.  
XX  
PD 21-MAR-2002.  
XX  
XX 10-SEP-2001; 2001WO-US026015.  
XX  
XX 11-SEP-2000; 2000US-00659671.  
PR  
XX (HYSE-) HYSEQ INC.  
PA  
XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;  
PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;  
PI  
XX WPI; 2002-292408/33.  
DR  
XX P-PSDB; ABB97229.  
DR  
XX  
PT An isolated polynucleotide for treating diseases associated with its  
PT encoded polypeptide such as cancer and multiple sclerosis.  
XX  
XX Claim 1; SEQ ID NO 53; 509pp; English.  
PS  
XX  
CC The present invention provides the protein and coding sequences of 444  
CC novel human proteins. These were isolated from expressed sequences tags  
CC (ESTs). They can be used to stimulate cell growth, to regulate  
CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth  
CC e.g. in burn treatment, to regulate the immune system e.g. to treat  
CC multiple sclerosis, to regulate activin or inhibit e.g. to treat  
CC infertility, to regulate haemostasis or thrombolysis e.g. to treat stroke

CC and cancer, to screen for drugs, to treat inflammatory conditions e.g.  
CC rheumatoid arthritis, and to treat nervous system disorders e.g.  
CC Parkinson's disease. The present sequence is a coding sequence of the  
CC invention  
XX  
SQ Sequence 1648 BP; 494 A; 365 C; 429 G; 357 T; 0 U; 3 Other;  
Query Match 50.0%; Score 21; DB 6; Length 1648;  
Best Local Similarity 73.0%; Pred. No. 1.4e+02;  
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
QY 5 ACGCTTGGCCGAGCCCTTACAAAGACTATGAAAGTAAA 41  
DB 624 AAGTTTCACAGAGTCCTTCAAAAGACTCTGGAAGAAA 660  
RESULT 89  
ADL62578  
ID ADL62578 standard; DNA; 2145 BP.  
XX  
AC ADL62578;  
XX  
XX 20-MAY-2004 (first entry)  
XX  
DE Human ovarian cancer DNA marker #20790.  
XX  
XX Human; ovarian cancer; ds; tumour; cytostatic; DNA marker.  
KW  
OS Homo sapiens.  
XX  
PN WO200170979-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 21-MAR-2001; 2001WO-US009126.  
XX  
PR 21-MAR-2000; 2000US-0191031P.  
PR 25-MAY-2000; 2000US-0207124P.  
PR 15-JUN-2000; 2000US-0211940P.  
PR 07-JUL-2000; 2000US-0216820P.  
PR 25-JUL-2000; 2000US-0220661P.  
PR 21-DEC-2000; 2000US-0257672P.  
XX  
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX  
XX Lee J, Lillie J;  
PI  
DR WPI; 2001-611502/70.  
XX  
PT Novel isolated nucleic acid molecules (markers) overexpressed in ovarian  
PT cancer cells as compared to their normal non-cancerous ovarian cells are  
PT used to characterize stage, grade, histological type of ovarian cancer.  
XX  
PS Disclosure; SEQ ID NO 20790; 106pp; English.  
XX  
CC The invention relates to nucleic acid markers which are overexpressed in  
CC ovarian cancer cells as compared to their expression in normal (i.e. non-  
CC cancerous) ovarian cells. The invention also relates to polypeptides  
CC encoded by the markers, antibodies that selectively bind to the  
CC polypeptides, a method of inhibiting ovarian cancer in a patient at risk  
CC of developing ovarian cancer involving inhibiting expression of a gene  
CC corresponding to a marker of the invention and a method of treating a  
CC patient afflicted with ovarian cancer comprising providing to cells of  
CC the patient an antisense oligonucleotide complementary to a marker of the  
CC invention. The markers are useful for assessing if a patient is afflicted  
CC with ovarian cancer, which involves comparing the level of expression of  
CC a marker in a patient sample and a normal level of expression of the  
CC marker in a control non-ovarian cancer sample. A difference between the  
CC expression levels indicates ovarian cancer. The level of expression of a  
CC marker corresponds to a secreted protein or to a transcribed  
CC polynucleotide or its portion. The level of expression of the marker is  
CC assessed by detecting the presence in the sample, a protein or protein  
CC fragment corresponding to the marker. The presence of protein or protein

CC fragment is detected using an antibody that specifically binds with the  
CC protein or protein fragment. Alternatively, the level of expression of  
CC the marker is assessed by detecting the presence of a transcribed  
CC polynucleotide which anneals with the marker or anneals with a portion of  
CC the polynucleotide comprising the marker, under stringent conditions. The  
CC marker is also used for monitoring the progression of ovarian cancer in a  
CC patient which involves detecting expression of the marker in a patient  
CC sample at a first point in time, repeating the method at a subsequent  
CC time and comparing the level of expression. The method is carried out  
CC using an ovarian tissue sample. A composition comprising a marker,  
CC polypeptide or antibody of the invention is used to treat ovarian cancer.  
CC This sequence represents a human ovarian cancer DNA marker of the  
CC invention.  
XX  
SQ Sequence 2145 BP; 595 A; 496 C; 577 G; 470 T; 0 U; 7 Other;  
Query Match 50.0%; Score 21; DB 5; Length 2145;  
Best Local Similarity 73.0%; Pred. No. 1.5e+02;  
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
QY 5 ACGCTTGGCCGAGCCCTTACAAAGACTATGAAAGTAAA 41  
DB 657 AAGTTTCACAGAGTCCTTCAAAAGACTCTGGAAGAAA 693  
RESULT 90  
ADS11553  
ID ADS11553 standard; DNA; 3870 BP.  
XX  
AC ADS11553;  
XX  
XX 16-DEC-2004 (first entry)  
XX  
DE Human therapeutic contig DNA - SEQ ID 1790.  
XX  
KW antiinflammatory; neuroprotective; antianaemic; cytostatic; vulnery;  
KW inflammatory; haematopoiesis; immunity; neurodegenerative; stem cell;  
KW aplastic anaemia; cancer; wound healing; gene therapy; ds; gene.  
OS Homo sapiens.  
XX  
PN WO2004080148-A2.  
XX  
PD 23-SEP-2004.  
XX  
PF 30-SEP-2003; 2003WO-US030720.  
XX  
PR 02-OCT-2002; 2002US-0416186P.  
XX  
PA (NUVE-) NUVELO INC.  
XX  
PI Tang YT, Asundi V, Ren F, Zhang J, Zhang J, Wehrman T, Wang Z, Ma Y;  
PI Wang D, Chen R, Zhao QA, Wang J, Ghosh M, Xue AJ, Weng G, Zhou P;  
XX  
XX WPI; 2004-669857/65.  
DR  
DR P-PSDB; ADS12151.  
XX  
XX New polynucleotide, useful in preparing a composition for diagnosing or  
PT treating inflammatory, neurodegenerative or stem cell disorders, e.g.,  
PT aplastic anemia or cancer for promoting wound healing.  
XX  
PS Example 2; SEQ ID NO 1790; 718pp; English.  
XX  
CC The invention relates to a novel isolated polynucleotide and the encoded  
CC polypeptide. The molecules of the invention demonstrate antiinflammatory,  
CC neuroprotective, antianaemic, cytostatic and vulnery activities and may  
CC be useful in preparing a composition for diagnosing or treating  
CC inflammatory, haematopoietic, immune, neurodegenerative or stem cell  
CC disorders, such as aplastic anaemia or cancer, as well as for promoting  
CC wound healing. The molecules may also be utilised during gene therapy  
CC procedures. The current sequence is that of a human therapeutic contig  
CC DNA of the invention. The current sequence is not shown explicitly within  
CC the specification but can be accessed from the WIPO web-site.

```
XX
SQ Sequence 3870 BP; 1071 A; 924 C; 938 G; 937 T; 0 U; 0 Other;

Query Match          50.0%; Score 21; DB 13; Length 3870;
Best Local Similarity 73.0%; Pred. No. 1.7e+02;
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 5 ACGCTTGGCGGAGCCCTTACAAAGACTATGAACTATAA 41
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 610 AAGTTTCACAGAGTCCTTCAAAAGACTCTGAAGAAAA 646

RESULT 91
ADI22610/c
ID ADI22610 standard; DNA; 4605 BP.
XX
AC ADI22610;
XX
DT 22-APR-2004 (first entry)
XX
DE Human liver differentially expressed cDNA seq id 420.
XX
KW hepatotropic; antiinflammatory; virucide; cytostatic; gene therapy;
KW differentially gene expression; liver; toxin; liver disorder;
KW biliary cirrhosis; X-linked adrenoleukodystrophy; Zellweger syndrome;
KW hepatorenal syndrome; hepatitis; hepatocarcinoma; metabolic response;
KW toxicological response; ss; EST; expressed sequence tag; human.
XX
OS Homo sapiens.
XX
PN US2003165854-A1.
XX
PD 04-SEP-2003.
XX
PF 05-DEC-2001; 2001US-00006285.
XX
PR 05-DEC-2000; 2000US-0251986P.
XX
PA (CUNN/) CUNNINGHAM M J.
PA (KASE/) KASER M R.
XX
PI Cunningham MJ, Kaser MR;
XX
DR WPI; 2003-863697/80.
XX
PT New combination comprising a number of cDNAs that are differentially
PT expressed in a liver treated with a toxin, useful for diagnosing, staging
PT or treating liver disorders (e.g. biliary cirrhosis, hepatitis or
PT hepatocarcinoma).
XX
PS Claim 1; SEQ ID NO 420; 28pp; English.
XX
CC The invention describes a combination comprising a number of cDNAs that
CC are differentially expressed in a liver treated with a toxin and are
CC selected from any of the 514 cDNAs listed in the specification, or their
CC complements. The combination is useful in diagnosing, staging or treating
CC a liver disorder (e.g. biliary cirrhosis, X-linked adrenoleukodystrophy,
CC Zellweger syndrome, hepatorenal syndrome, hepatitis or hepatocarcinoma),
CC in monitoring diagnostic and therapeutic applications, in detecting
CC metabolic and toxicological responses, and in elucidating drug mechanism
CC of action. This sequence represents a cDNA differentially expressed in
CC liver tissues in response to treatment with a toxin.
XX
SQ Sequence 4605 BP; 1065 A; 1206 C; 1133 G; 1201 T; 0 U; 0 Other;

Query Match          50.0%; Score 21; DB 10; Length 4605;
Best Local Similarity 73.0%; Pred. No. 1.8e+02;
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 5 ACGCTTGGCGGAGCCCTTACAAAGACTATGAACTATAA 41
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3959 AAGTTTCACAGAGTCCTTCAAAAGACTCTGAAGAAAA 3923

us-10-600-816-30.rng

RESULT 92
ADQ22756
ID ADQ22756 standard; DNA; 6346 BP.
XX
AC ADQ22756;
XX
DT 26-AUG-2004 (first entry)
XX
DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 5576.
XX
KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
KW ds.
XX
OS Homo sapiens.
XX
PN WO2004048938-A2.
XX
PD 10-JUN-2004.
XX
PF 26-NOV-2003; 2003WO-US038193.
XX
PR 26-NOV-2002; 2002US-0429739P.
XX
PA (PROT-) PROTEIN DESIGN LABS INC.
XX
PI Aziz N, Ginsburg WM, Zlotnik A;
XX
DR WPI; 2004-441208/41.
XX
PT Early detection of soft tissue sarcoma comprises determining expression
PT of a gene in a first soft tissue sample and a normal soft tissue sample
PT and comparing the gene expression, also useful in treating soft tissue
PT sarcoma.
XX
PS Example 2; SEQ ID NO 5576; 210pp; English.
XX
CC The invention relates to a novel method for detecting soft tissue sarcoma
CC which comprises obtaining a first soft tissue sample from an individual
CC and a normal soft tissue sample from the same or different individual,
CC determining the expression of a gene in both samples and comparing the
CC expression of the gene in both soft tissue samples, where a higher level
CC of protein expression in the first soft tissue sample indicates the
CC presence of soft tissue sarcoma. The method of the invention has
CC cytostatic applications and may be useful for detecting soft tissue
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC acid sequences may be useful in diagnostic and screening applications.
CC The current sequence is that of a human soft tissue sarcoma-upregulated
CC DNA of the invention. The current sequence is not shown within the
CC specification per se but was submitted in CD format by the inventor.
XX
SQ Sequence 6346 BP; 1533 A; 1636 C; 1775 G; 1351 T; 0 U; 51 Other;

Query Match          50.0%; Score 21; DB 12; Length 6346;
Best Local Similarity 73.0%; Pred. No. 1.9e+02;
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 5 ACGCTTGGCGGAGCCCTTACAAAGACTATGAACTATAA 41
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 659 AAGTTTCACAGAGTCCTTCAAAAGACTCTGAAGAAAA 695

RESULT 93
ADC37602
ID ADC37602 standard; DNA; 7174 BP.
XX
AC ADC37602;
XX
DT 18-DEC-2003 (first entry)
XX
DE Human nucleic acid associated protein NAAP-9 coding sequence.
XX
KW Human; nucleic acid associated protein; NAAP; cytostatic;
```

```
KW antiarteriosclerotic; anticonvulsant; nootropic; neuroprotective;
KW cerebroprotective; anti-HIV; anti-allergic; anti-inflammatory;
KW thymometric; gene therapy; cell proliferative disorder; cancer;
KW atherosclerosis; neurological disorder; epilepsy; Huntington's disease;
KW stroke; immune disorder; inflammatory disorder; AIDS; allergy;
KW developmental disorder; Hypothyroidism; Cushing's syndrome; infection;
KW gene; db.
XX
XX Homo sapiens.
XX
XX WO2003046151-A2.
XX
XX 05-JUN-2003.
XX
XX 26-NOV-2002; 2002WO-US038445.
XX
XX 27-NOV-2001; 2001US-0333925P.
XX
XX 07-DEC-2001; 2001US-0340477P.
XX
XX 14-DEC-2001; 2001US-0340362P.
XX
XX 18-DEC-2001; 2001US-0342002P.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Baughn MR, Becha SD, Bhatia U, Blake JJ, Burford N, Burrill JD,
XX Chang H, Chawla NK, Elliott VS, Emerling BM, Forsythe IJ, Gandhi AR,
XX Gietzen KJ, Gorvad AE, Griffin JA, Hafalia AUA, Jackson JL, Ho A,
XX Ison CH, Jackson AA, Jiang X, Jin P, Kable AE, Khare R, Lal PG,
XX Lee EA, Lee S, Lee SY, Li JX, Lu DAM, Ramkumar J, Richardson TW,
XX Sprague WW, Swarnakar A, Tang YT, Warren BA, Xu Y, Yao MG, Yue H,
XX Zheng W;
XX WPI; 2003-513642/48.
XX
XX P-PSDB; ADC37542.
XX
XX New human nucleic acid associated proteins (NAAP), useful for diagnosing,
XX treating and preventing diseases or conditions associated with the
XX aberrant NAAP expression e.g. cancer, AIDS, atherosclerosis, epilepsy, or
XX infections.
XX
XX Claim 5; SEQ ID NO 69; 383pp; English.
XX
XX The present invention relates to novel human nucleic acid associated
XX protein (NAAP) (I; ADC37534-ADC37593) and their coding sequences ({}). The
XX NAAPs and their coding sequences are useful in diagnosing, treating and
XX preventing diseases or conditions associated with the decreased
XX expression or over expression of NAAP, such as cell proliferative (e.g.
XX cancer, atherosclerosis), neurological (e.g. epilepsy, Huntington's
XX disease, stroke), immune/inflammatory (e.g. AIDS, allergies) and
XX developmental (e.g. Hypothyroidism, Cushing's syndrome) disorders, or
XX infections. These are also useful in assessing the effects of exogenous
XX compounds on the expression of nucleic acid and amino acid sequences of
XX NAAP.
XX
XX Sequence 7174 BP; 1812 A; 1991 C; 1933 G; 1438 T; 0 U; 0 Other;
XX
XX Query Match 50.0%; Score 21; DB 10; Length 7174;
XX Best Local Similarity 73.0%; Pred. No. 2e+02;
XX Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
XX
XX QY 6 CGCTTGGCGGAGCCCTTACAAAGACTATGAGTAAAG 42
XX
XX 460 CTCGTGGCGGAGCGATCACAATTACAAATGCAAGTAAAG 496
XX
XX
XX RESULT 94
XX ADA02708/c
XX ID ADA02708 standard; DNA; 96589 BP.
XX
XX AC ADA02708;
XX
XX DT 06-NOV-2003 (first entry)
XX
XX KW Human ZFXH1B carcinoma associated gene, SEQ ID NO:1226.
XX
XX DE
```

---

```
XX
XX Human; carcinoma associated; oncogene; carcinoma; cancer; breast;
XX prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;
XX gene; ds.
XX
XX Homo sapiens.
XX
XX WO2003057146-A2.
XX
XX 17-JUL-2003.
XX
XX 26-DEC-2002; 2002WO-US041414.
XX
XX 26-DEC-2001; 2001US-00035832.
XX
XX (SAGR-) SAGRES DISCOVERY.
XX
XX Morris DW;
XX
XX WPI; 2003-587068/55.
XX
XX New recombinant nucleic acid encoding carcinoma associated protein,
XX useful for preparing compositions for treating carcinomas.
XX
XX Claim 1; SEQ ID NO 1226; 245pp; English.
XX
XX The invention relates to recombinant carcinoma associated (CA) nucleic
XX acid sequences from mouse and human (ADA01482-ADA03094), and to
XX recombinant carcinoma associated proteins (CAP) encoded by them. The
XX invention also encompasses expression vectors and host cells comprising a
XX CA nucleic acid, a polypeptide (especially an antibody) that specifically
XX binds to the protein, and a biochip comprising CA nucleic acid or
XX fragments thereof. The sequences of the invention were identified using
XX oncogenic retroviruses, which insert into the genome of the host organism
XX at random. Many of these do not carry transduced host oncogenes or
XX pathogenic trans-acting viral genes, meaning that cancer incidence is a
XX direct consequence of the effects of proviral integration into host
XX protooncogenes. The CA nucleic acid sequences can be used to diagnose
XX carcinoma (especially breast cancer, prostate cancer, lymphoma or
XX leukaemia) or a propensity to carcinoma by determination of the sequence
XX of a CA gene, or by determination of CA gene expression in particular
XX tissues. CA nucleic acids, proteins and antibodies are also useful as
XX therapeutic agents and in screening and evaluating drug candidates. The
XX present sequence represents a specifically claimed human CA nucleic acid
XX sequence of the invention. Note: The complete sequence data for this
XX patent did not form part of the printed specification, but was obtained
XX in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 96589 BP; 29004 A; 18509 C; 19494 G; 29562 T; 0 U; 20 Other;
XX
XX Query Match 50.0%; Score 21; DB 9; Length 96589;
XX Best Local Similarity 73.0%; Pred. No. 3.5e+02;
XX Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
XX
XX QY 5 AGCTTGGCGGAGCCCTTACAAAGACTATGAGTAAAG 41
XX
XX 45448 ACTCTCTGCCGCGCTGTACAGAAATTATGAGAAAA 45412
XX
XX
XX RESULT 95
XX ADB72446/c
XX ID ADB72446 standard; DNA; 96589 BP.
XX
XX AC ADB72446;
XX
XX DT 04-DEC-2003 (first entry)
XX
XX DE Human ZFXH1B gene.
XX
XX KW human; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;
XX cancer; neoplasm; adenocarcinoma; sarcoma; gene.
XX
XX
```







Query Match 49.0%; Score 20.6; DB 9; Length 457;  
Best Local Similarity 85.2%; Pred. No. 1.6e+02;  
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 15 GAGCCCTTACAAAGACTATGAAGTAAA 41  
|||||  
Db 101 GAGCCCTTACAAAGACTATGAAGCACCANA 75

RESULT 100  
ADL45933  
ID ADL45933 standard; DNA; 682 BP.  
XX  
AC ADL45933;  
XX  
DT 20-MAY-2004 (first entry)  
XX  
XX Human ovarian cancer DNA marker #19823.

XX Human; ovarian cancer; ds; tumour; cytostatic; DNA marker.  
XX  
XX Homo sapiens.  
XX WO200170979-A2.  
XX  
XX 27-SEP-2001.  
XX  
XX 21-MAR-2001; 2001WO-US0009126.  
XX  
XX 21-MAR-2000; 2000US-0191031P.  
XX  
XX 25-MAY-2000; 2000US-0207124P.  
XX  
XX 15-JUN-2000; 2000US-0211940P.  
XX  
XX 07-JUL-2000; 2000US-0216820P.  
XX  
XX 25-JUL-2000; 2000US-0220661P.  
XX  
XX 21-DEC-2000; 2000US-0257672P.  
XX  
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX Lee J, Lillie J;  
XX WPI; 2001-611502/70.  
XX  
XX Novel isolated nucleic acid molecules (markers) overexpressed in ovarian  
XX cancer cells as compared to their normal non-cancerous ovarian cells are  
XX used to characterize stage, grade, histological type of ovarian cancer.  
XX  
XX Disclosure; SEQ ID NO 19823; 106pp; English.

XX The invention relates to nucleic acid markers which are overexpressed in  
XX ovarian cancer cells as compared to their expression in normal (i.e. non-  
XX cancerous) ovarian cells. The invention also relates to polypeptides  
XX encoded by the markers, antibodies that selectively bind to the  
XX polypeptides, a method of inhibiting ovarian cancer in a patient at risk  
XX of developing ovarian cancer involving inhibiting expression of a gene  
XX corresponding to a marker of the invention and a method of treating a  
XX patient afflicted with ovarian cancer comprising providing to cells of  
XX the patient an antisense oligonucleotide complementary to a marker of the  
XX invention. The markers are useful for assessing if a patient is afflicted  
XX with ovarian cancer, which involves comparing the level of expression of  
XX a marker in a patient sample and a normal level of expression of the  
XX marker in a control non-ovarian cancer sample. A difference between the  
XX expression levels indicates ovarian cancer. The level of expression of a  
XX marker corresponds to a secreted protein or to a transcribed  
XX polynucleotide or its portion. The level of expression of the marker is  
XX assessed by detecting the presence in the sample; a protein or protein  
XX fragment corresponding to the marker. The presence of protein or protein  
XX fragment is detected using an antibody that specifically binds with the  
XX protein or protein fragment. Alternatively, the level of expression of  
XX polynucleotide which anneals with the marker or anneals with a portion of  
XX the polynucleotide comprising the marker, under stringent conditions. The  
XX marker is also used for monitoring the progression of ovarian cancer in a

CC patient which involves detecting expression of the marker in a patient  
CC sample at a first point in time, repeating the method at a subsequent  
CC time and comparing the level of expression. The method is carried out  
CC using an ovarian tissue sample. A composition comprising a marker,  
CC polypeptide or antibody of the invention is used to treat ovarian cancer.  
CC This sequence represents a human ovarian cancer DNA marker of the  
CC invention. Note: The sequence data for this patent did not form part of  
CC the printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX  
SQ Sequence 682 BP; 235 A; 112 C; 115 G; 209 T; 0 U; 11 Other;

Query Match 49.0%; Score 20.6; DB 5; Length 682;  
Best Local Similarity 85.2%; Pred. No. 1.7e+02;  
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 15 GAGCCCTTACAAAGACTATGAAGTAAA 41  
|||||  
Db 436 GAGCCCTTACAAAGACTATGAAGCACCANA 462

Search completed: June 10, 2006, 15:32:20  
Job time : 351.45 secs

**THIS PAGE IS BLANK**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2006, 15:19:28 ; Search time 5182.8 Seconds  
(without alignments)  
453.155 Million cell updates/sec

Title: US-10-600-816-30

Perfect score: 42

Sequence: 1 gccacgctggccgagcc.....acaaagactatgaagtaag 42

Scoring table: IDENTITY\_NUC

Gapop 10:0 , Gapext 1.0

Searched: 48236798 seqs, 27959665780 residues

Total number of hits satisfying chosen parameters: 96473596

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

Database :

EST:\*

1: gb\_est1:\*

2: gb\_est3:\*

3: gb\_est4:\*

4: gb\_est5:\*

5: gb\_est6:\*

6: gb\_est7:\*

7: gb\_est8:\*

8: gb\_est9:\*

9: gb\_est10:\*

10: gb\_gss1:\*

11: gb\_gss2:\*

12: gb\_gss3:\*

13: gb\_gss4:\*

14: gb\_gss5:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42	100.0	158	2	B1063311 IL3-UT011
C 2	42	100.0	307	7	B8818833 CM2-BN030
C 3	42	100.0	341	7	B8818817 CM2-BN030
C 4	42	100.0	384	7	B8818835 CM2-BN030
C 5	42	100.0	396	7	B8818828 CM2-BN030
C 6	42	100.0	404	7	B8818840 CM2-BN030
C 7	42	100.0	424	7	B8818805 CM2-BN030
C 8	42	100.0	453	5	CD672930 fg18a05.Y
C 9	42	100.0	458	7	B8818871 CM2-BN030
C 10	42	100.0	466	7	B8818859 CM2-BN030
C 11	42	100.0	467	7	B8818852 CM2-BN030
C 12	42	100.0	571	3	BP253378 BP253378
C 13	42	100.0	577	4	CB152636 K-EST0209
C 14	42	100.0	580	3	BP265234 BP265234
C 15	42	100.0	580	3	BP268151 BP268151
C 16	42	100.0	582	3	BP256144 BP256144
C 17	42	100.0	582	3	BP363758 BP363758
C 18	42	100.0	584	3	BP256147 BP256147
C 19	42	100.0	621	4	CB130950 K-EST0180

BE747107	601580680	7	BE747107	630	42	100.0
BM843051	K-EST0120	637	BM843051	637	42	100.0
BF568108	602183908	699	BF568108	699	42	100.0
BG251131	602164932	702	BG251131	702	42	100.0
BQ887352	AGENCOURT	851	BQ887352	851	42	100.0
CA454746	AGENCOURT	865	CA454746	865	42	100.0
BM011054	603634744	877	BM011054	877	42	100.0
BQ958341	AGENCOURT	899	BQ958341	899	42	100.0
BG831564	602765819	901	BG831564	901	42	100.0
BX372160	BX372160	918	BX372160	918	42	100.0
BU543952	AGENCOURT	923	BU543952	923	42	100.0
BX370558	BX370558	962	BX370558	962	42	100.0
BQ067433	AGENCOURT	1023	BQ067433	1023	42	100.0
CR597125	full-length	1830	CR597125	1830	42	100.0
CF780868	AGENCOURT	916	CF780868	916	41	97.6
B1193620	602946519	1125	B1193620	1125	41	97.6
BE818831	CM2-BN030	365	BE818831	365	40.4	96.2
BE818807	CM2-BN030	449	BE818807	449	40.4	96.2
BP267627	BP267627	580	BP267627	580	40.4	96.2
BQ685729	AGENCOURT	940	BQ685729	940	40.4	96.2
AU140676	AU140676	680	AU140676	680	33.6	80.0
AA112374	zn68e10.r	367	AA112374	367	31	73.8
BE006062	RC2-BN012	583	BE006062	583	30.4	72.4
DN875318	nad38g08.	606	DN875318	606	29.2	69.5
BE548141	601072396	912	BE548141	912	28.4	67.6
AW390096	CM3-ST017	528	AW390096	528	27	64.3
BP267918	BP267918	582	BP267918	582	27	64.3
BE815636	PM4-BN017	240	BE815636	240	26.2	62.4
AA222951	mw01h08.r	336	AA222951	336	24.2	57.6
W10531	ma42b12.r1	49	W10531	49	24.2	57.6
BE003240	QV4-BN009	374	BE003240	374	23.2	55.2
AL241805	Tetraodon	907	AL241805	907	23.2	55.2
BP267556	BP267556	578	BP267556	578	23	54.8
C45442	C45442 Yuji	360	C45442	360	22.8	54.3
CE847883	ti9r-ges-	630	CE847883	630	22.8	54.3
CK976767	4107795.B	715	CK976767	715	22.8	54.3
DN020621	JGI_CAAAR2	864	DN020621	864	22.8	54.3
AA615518	vo7e02.r	454	AA615518	454	22.6	53.8
CB818826	EST 1716	754	CB818826	754	22.6	53.8
DV624038	93529.1.C	609	DV624038	609	22.4	53.3
AQ855455	cp61823A	746	AQ855455	746	22.4	53.3
CK718467	18536.Swo	765	CK718467	765	22.4	53.3
CV286484	66768.1.A	810	CV286484	810	22.4	53.3
BE004395	CM0-BN010	338	BE004395	338	22.2	52.9
AA956676	UI-R-E1-f	362	AA956676	362	22.2	52.9
CB727988	AMGNNUC.U	482	CB727988	482	22.2	52.9
AI171672	EST217647	484	AI171672	484	22.2	52.9
BP470636	BP470636	498	BP470636	498	22.2	52.9
AI175079	EST218597	515	AI175079	515	22.2	52.9
AI113032	UI-R-E1-g	526	AI113032	526	22.2	52.9
BE116174	UI-R-BSI-	532	BE116174	532	22.2	52.9
AI170330	EST216256	536	AI170330	536	22.2	52.9
CF978837	FAR004B12	545	CF978837	545	22.2	52.9
BP499326	BP499326	552	BP499326	552	22.2	52.9
AA818700	UI-R-A0-a	584	AA818700	584	22.2	52.9
AI599233	EST250936	612	AI599233	612	22.2	52.9
BM386471	UI-R-CM1-	614	BM386471	614	22.2	52.9
AA946356	EST201855	620	AA946356	620	22.2	52.9
CB326875	UI-R-F50-	673	CB326875	673	22.2	52.9
CA504656	UI-R-FJ0-	689	CA504656	689	22.2	52.9
CK838771	UI-R-AC1-	736	CK838771	736	22.2	52.9
CZ983576	197310.T0	859	CZ983576	859	22.2	52.9
CO754450	Mdft3050	114	CO754450	114	22	52.4
AV238654	AV238654	266	AV238654	266	22	52.4
CO755512	Mdft3049	381	CO755512	381	22	52.4
CO541025	Mdft3038	434	CO541025	434	22	52.4
CO416452	Mdft2065n	446	CO416452	446	22	52.4
BQ122319	EST607895	476	BQ122319	476	22	52.4
CV657943	Mdft-6016c	482	CV657943	482	22	52.4
CV084234	Mdft5027m	487	CV084234	487	22	52.4
CO415972	Mdft3009b	535	CO415972	535	22	52.4
DB182714	DB182714	554	DB182714	554	22	52.4
CV880385	Mdft6018b	567	CV880385	567	22	52.4

C 93	22	52.4	647	11	AZ995992	AZ995992 2M0282G06	166	21.2	50.5	759	5	CJ479497	CJ479497
C 94	22	52.4	655	8	CN916071	CN916071 030709ABO	C 167	21.2	50.5	787	4	CA457909	CA457909
C 95	22	52.4	683	8	CN900285	CN900285 031104ABA	168	21.2	50.5	794	5	CJ418153	CJ418153
C 96	22	52.4	703	8	C0398610	C0398610 AGENCOURT	169	21.2	50.5	816	10	DV520386	DV520386
C 97	22	52.4	745	2	B1561919	B1561919 603254822	170	21.2	50.5	844	8	CO448028	CO448028
C 98	22	52.4	747	11	BZ436965	BZ436965 BONEV455F	C 171	21.2	50.5	867	14	CT417138	CT417138
C 99	22	52.4	762	7	BF112154	BF112154 7148a08.x	172	21.2	50.5	874	8	CN156428	CN156428
C 100	22	52.4	771	11	BZ455062	BZ455062 BONGH48TR	C 173	21.2	50.5	902	8	CO452256	CO452256
C 101	22	52.4	840	14	DU847910	DU847910 32904 Tom	C 174	21.2	50.5	911	12	CC400755	CC400755
C 102	22	52.4	860	2	B1453797	B1453797 603175062	175	21.2	50.5	1043	5	CF585003	CF585003
C 103	21.8	51.9	478	2	BG037359	BG037359 dg49b05.y	176	21.2	50.5	2268	13	CL981290	CL981290
C 104	21.8	51.9	524	7	BF083559	BF083559 MR0-CT045	C 177	21	50.0	186	11	AZ3110055	AZ3110055
C 105	21.8	51.9	532	14	AG931679	AG931679 Drosophill	178	21	50.0	248	1	AA383715	AA383715
C 106	21.8	51.9	538	8	CV938373	CV938373 PMRpm0 0	179	21	50.0	264	1	AV283291	AV283291
C 107	21.8	51.9	609	2	BG961171	BG961171 PMO-CT082	C 180	21	50.0	296	7	AW327096	AW327096
C 108	21.8	51.9	656	4	BX845889	BX845889 BX845889	C 181	21	50.0	337	7	AW939011	AW939011
C 109	21.8	51.9	731	3	BU246088	BU246088 603593014	C 182	21	50.0	345	7	BE478490	BE478490
C 110	21.8	51.9	801	12	CC411501	CC411501 PUHME48TD	C 183	21	50.0	368	7	AW327024	AW327024
C 111	21.8	51.9	815	13	CL311198	CL311198 293257_LB	C 184	21	50.0	436	8	AJ647102	AJ647102
C 112	21.8	51.9	866	14	CT393924	CT393924 Sus scrofa	C 185	21	50.0	446	7	AW478519	AW478519
C 113	21.8	51.9	941	12	CG163630	CG163630 PUFOB15TB	C 186	21	50.0	469	3	BQ319886	BQ319886
C 114	21.6	51.4	170	3	BP726758	BP726758 BP726758	C 187	21	50.0	498	2	B1536748	B1536748
C 115	21.6	51.4	298	1	AA935101	AA935101 nv38g11.s	C 188	21	50.0	520	7	AW367992	AW367992
C 116	21.6	51.4	497	2	BG345699	BG345699 dg40a09.y	C 189	21	50.0	528	2	BG967089	BG967089
C 117	21.6	51.4	617	2	BJ636044	BJ636044 Buj636044	C 190	21	50.0	528	2	BM259003	BM259003
C 118	21.6	51.4	686	2	BJ643078	BJ643078 Buj643078	C 191	21	50.0	545	2	B1541850	B1541850
C 119	21.6	51.4	690	3	BP726506	BP726506 BP726506	C 192	21	50.0	549	7	AW632501	AW632501
C 120	21.6	51.4	715	2	BJ097745	BJ097745 BJ097745	C 193	21	50.0	551	2	BF725048	BF725048
C 121	21.6	51.4	1452	14	AJ861368	AJ861368 Brassica	194	21	50.0	551	10	T79981	T79981
C 122	21.4	51.0	329	11	AQ650989	AQ650989 Sheared D	195	21	50.0	582	3	BP309333	BP309333
C 123	21.4	51.0	343	12	CE485273	CE485273 tigr-gss-	196	21	50.0	602	7	AW953241	AW953241
C 124	21.4	51.0	391	7	AW771362	AW771362 hm55b11.x	197	21	50.0	627	1	AJ656152	AJ656152
C 125	21.4	51.0	432	1	AA483518	AA483518 ne40c04.s	198	21	50.0	627	1	BF305629	BF305629
C 126	21.4	51.0	443	3	B0830662	B0830662 T011D03 P	199	21	50.0	627	10	DV934823	DV934823
C 127	21.4	51.0	445	11	AQ092573	AQ092573 HS 3003_B	200	21	50.0	645	8	CN364607	CN364607
C 128	21.4	51.0	456	1	AJ789122	AJ789122 AJ789122_	201	21	50.0	647	8	CN348583	CN348583
C 129	21.4	51.0	528	7	AW152265	AW152265 xg40f10.x	202	21	50.0	649	9	DN276014	DN276014
C 130	21.4	51.0	556	3	B0866894	B0866894 S072A02 P	203	21	50.0	681	4	CB852045	CB852045
C 131	21.4	51.0	585	11	AQ651424	AQ651424 Sheared D	C 204	21	50.0	702	5	BE303650	BE303650
C 132	21.4	51.0	593	1	AJ560115	AJ560115 AJ560115	205	21	50.0	705	4	CB167757	CB167757
C 133	21.4	51.0	627	1	AJ559741	AJ559741 AJ559741	206	21	50.0	716	2	B1850471	B1850471
C 134	21.4	51.0	639	1	AJ800854	AJ800854 AJ800854	207	21	50.0	717	2	BI771708	BI771708
C 135	21.4	51.0	661	13	CW225615	CW225615 104_663_1	C 208	21	50.0	717	10	DT659125	DT659125
C 136	21.4	51.0	720	12	CE231950	CE231950 tigr-gss-	209	21	50.0	722	5	CO885662	CO885662
C 137	21.4	51.0	742	8	CX176408	CX176408 B11_69-59	210	21	50.0	729	9	DN823839	DN823839
C 138	21.4	51.0	818	14	DU480901	DU480901 109841593	C 211	21	50.0	729	9	CB951670	CB951670
C 139	21.4	51.0	877	10	DT479357	DT479357 WS02525.B	212	21	50.0	734	4	BG771856	BG771856
C 140	21.4	51.0	1071	14	CNS03P11	AC254710 Tetraodon	213	21	50.0	735	2	BG700459	BG700459
C 141	21.4	51.0	1105	12	CC233180	CC233180 CH261-2B1	214	21	50.0	764	2	B1828470	B1828470
C 142	21.4	51.0	1179	8	CX113082	CX113082 EI079N05	215	21	50.0	771	8	CN364605	CN364605
C 143	21.4	51.0	1234	12	CC293179	CC293179 CH261-2B1	216	21	50.0	772	3	BP442373	BP442373
C 144	21.4	51.0	218	5	CF045602	CF045602 QCK15f03.	217	21	50.0	774	5	CD628174	CD628174
C 145	21.2	50.5	306	4	BY356179	BY356179 BY356179	218	21	50.0	777	5	CJ038575	CJ038575
C 146	21.2	50.5	343	13	CZ072288	TFGA_SSR4	219	21	50.0	781	5	CD628175	CD628175
C 147	21.2	50.5	474	7	BB759691	BB759691 BB759691	C 220	21	50.0	783	2	B1766343	B1766343
C 148	21.2	50.5	514	11	BH767410	BH767410 EMBAC348H	221	21	50.0	785	2	BG323665	BG323665
C 149	21.2	50.5	516	11	BH083456	BH083456 RFC1-24-3	222	21	50.0	802	2	BI833072	BI833072
C 150	21.2	50.5	519	4	BY479172	BY479172 BY479172	C 223	21	50.0	808	14	CNS027WH	CNS027WH
C 151	21.2	50.5	534	11	AZ373445	AZ373445 1M0068E07	224	21	50.0	817	2	BG786507	BG786507
C 152	21.2	50.5	538	14	DE110180	DE110180 Oryzias 1	225	21	50.0	820	3	BP433505	BP433505
C 153	21.2	50.5	540	14	TA328C03Q	TA328C03Q T. brucei	226	21	50.0	844	2	BI114025	BI114025
C 154	21.2	50.5	545	5	CF002812	CF002812 QBH16B07.	227	21	50.0	850	2	BI666483	BI666483
C 155	21.2	50.5	548	4	CA591375	CA591375 hab62E02.	C 228	21	50.0	852	3	BP171826	BP171826
C 156	21.2	50.5	560	12	CC962357	CC962357 B0ICE13TR	229	21	50.0	857	13	CZ545391	CZ545391
C 157	21.2	50.5	562	11	AQ026479	AQ026479 HS 3241_A	230	21	50.0	869	4	BQ684368	BQ684368
C 158	21.2	50.5	575	14	CR053610	CR053610 Forward_s	231	21	50.0	884	3	BQ684368	BQ684368
C 159	21.2	50.5	583	3	BF218119	BF218119 BP218119	232	21	50.0	894	4	BX340888	BX340888
C 160	21.2	50.5	608	4	CA589246	CA589246 hab55f11D.	233	21	50.0	906	7	BF307808	BF307808
C 161	21.2	50.5	650	5	CD093738	CD093738 G356.111D	234	21	50.0				
C 162	21.2	50.5	673	14	DU472242	DU472242 109841568	235	21	50.0				
C 163	21.2	50.5	720	9	DN211676	DN211676 MEST928 D	236	21	50.0				
C 164	21.2	50.5	741	9	DN213756	DN213756 MEST986 D	237	21	50.0				
C 165	21.2	50.5	745	4	CB571386	CB571386 AGENCOURT	238	21	50.0				

239	21	50.0	912	3	BQ941302	BQ941302 AGENCOURT	312	20.6	49.0	302	8	CN348582	170005328
240	21	50.0	916	3	BQ523575	BQ523575 AGENCOURT	313	20.6	49.0	318	1	AI766278	hr71a08.x
241	21	50.0	935	2	BG118825	BG118825 602347821	314	20.6	49.0	325	7	BE221862	wh60e04.x
242	21	50.0	945	3	BQ677362	BQ677362 AGENCOURT	C 315	20.6	49.0	327	10	R72724	vj95d07.r1
243	21	50.0	959	4	BK402991	BK402991 EX402991	C 316	20.6	49.0	331	7	BE858505	7937e12.x
244	21	50.0	971	2	BG028292	BG028292 602295531	C 317	20.6	49.0	336	1	AI914938	wf81a07.x
245	21	50.0	1008	14	AY418464	AY418464 Homo sapi	C 318	20.6	49.0	339	1	AA234147	zr50d01.r
246	21	50.0	1008	14	AY418465	AY418465 Pan trogl	C 319	20.6	49.0	340	1	AV657285	AV657285
C 247	21	50.0	1028	4	BX386750	BX386750 BX386750	320	20.6	49.0	342	8	CN388117	170006008
248	21	50.0	1034	4	BK439606	BK439606 BX439606	321	20.6	49.0	348	1	AI373517	qz46f04.x
249	21	50.0	1050	2	BG028965	BG028965 602293092	322	20.6	49.0	348	1	AA547991	nj39g01.s
250	21	50.0	1051	2	BM478232	BM478232 AGENCOURT	C 323	20.6	49.0	357	1	AA733193	zg78h03.s
251	21	50.0	1085	5	CK209334	CK209334 FGAS02109	C 324	20.6	49.0	366	10	N62066	NG2066 EST58h085.B
C 252	21	50.0	1101	14	CNS00LTT	AL078741 Drosophi1	325	20.6	49.0	370	10	N90001	2b37a01.s1
C 253	21	50.0	1121	1	AL570641	AL570641 AL570641	326	20.6	49.0	382	7	BF088338	CM1-HT087
254	21	50.0	1141	3	BM925283	BM925283 AGENCOURT	327	20.6	49.0	382	7	BF155537	CM1-HT087
255	21	50.0	1545	6	CR625633	CR625633 full-leng	C 328	20.6	49.0	384	4	BY669557	BY669557
256	21	50.0	1586	6	BC071980	BC071980 Homo sapi	C 329	20.6	49.0	384	12	CE369298	CE369298 tigr-gss-
257	21	50.0	1628	6	CR609491	CR609491 full-leng	C 330	20.6	49.0	389	8	CN654263	CN654263 UMC-bcl.0
258	20.8	49.5	314	10	H83855	H83855 ys92f09.s1	C 331	20.6	49.0	390	11	BH043048	BH043048 RPCI-21-2
C 259	20.8	49.5	323	9	DN774646	DN774646 E2521 [C8	332	20.6	49.0	391	1	AA921695	AA921695 ol90e06.s
C 260	20.8	49.5	342	12	CG051017	CG051017 PUJAW63TB	333	20.6	49.0	391	1	AI022572	AI022572 ow46g05.s
261	20.8	49.5	346	12	CG051019	CG051019 PUJAW63TD	C 334	20.6	49.0	393	2	BM434423	BM434423 1RT10D12
C 262	20.8	49.5	388	1	AI653532	AI653532 tq95f03.x	335	20.6	49.0	393	7	BF449065	BF449065 7064g02.x
C 263	20.8	49.5	411	1	AI707946	AI707946 as34g12.x	C 336	20.6	49.0	402	7	BB669069	BB669069 BB669069
264	20.8	49.5	417	5	CJ222441	CJ222441 CJ222441	C 337	20.6	49.0	406	7	BB685308	BB685308 BB685308
265	20.8	49.5	478	8	CO818005	CO818005 FA-SEA001	C 338	20.6	49.0	406	8	CN348581	170004553
266	20.8	49.5	483	3	BQ105696	BQ105696 fc0020.e	C 339	20.6	49.0	409	1	AI219500	AI219500 qh24b03.x
267	20.8	49.5	512	7	BE234005	BE234005 140576.MA	C 340	20.6	49.0	420	1	AA812744	AA812744 aj43f04.s
268	20.8	49.5	517	7	BF644767	BF644767 NF051A12E	C 341	20.6	49.0	431	1	AI373151	AI373151 qz98h01.x
C 269	20.8	49.5	515	1	AI768299	AI768299 wg83g07.x	342	20.6	49.0	431	11	BZ325505	BZ325505 ic53h09.g
C 270	20.8	49.5	540	2	BI340991	BI340991 368174.MA	343	20.6	49.0	432	1	AI862299	AI862299 tw71e02.x
C 271	20.8	49.5	553	9	DN747454	DN747454 E4750 [C8	344	20.6	49.0	432	1	AI862299	AI862299 tw71e02.x
C 272	20.8	49.5	555	10	DM470330	DM470330 SGP311621	C 345	20.6	49.0	432	3	BQ030241	BQ030241 UI-H-DT0
C 273	20.8	49.5	559	11	BZ095512	BZ095512 CT230-215	C 346	20.6	49.0	433	1	AA460864	AA460864 zx69b06.r
C 274	20.8	49.5	567	14	DR37B12S	AI974341 Danilo rer	C 347	20.6	49.0	434	2	BG956844	BG956844 PM4-CT080
C 275	20.8	49.5	571	2	BI342543	BI342543 370375.MA	C 348	20.6	49.0	435	4	BY553103	BY553103 BY553103
C 276	20.8	49.5	579	1	AI719565	AI719565 as54d11.x	C 349	20.6	49.0	436	1	AW137818	AW137818 UI-H-B11-
C 277	20.8	49.5	592	8	CX001150	CX001150 iv39d06.b	C 350	20.6	49.0	438	1	AA977781	AA977781 oq59h07.s
C 278	20.8	49.5	602	3	BUE94039	BUE94039 EST0076.A	351	20.6	49.0	439	1	AA740912	AA740912 nz02g10.s
C 279	20.8	49.5	605	14	CR62794	CR62794 mth4-12J6	352	20.6	49.0	439	1	BE219295	BE219295 hv5a111.x
C 280	20.8	49.5	635	9	CX631471	CX631471 GAN002E01	353	20.6	49.0	440	1	AI375147	AI375147 tc09h10.x
C 281	20.8	49.5	639	12	CG735356	CG735356 GUAV34TH	354	20.6	49.0	444	1	AI014659	AI014659 ovi34f12.x
C 282	20.8	49.5	655	13	CW283279	CW283279 104_759.1	355	20.6	49.0	444	1	AI474043	AI474043 ti69f01.x
283	20.8	49.5	674	13	CW283276	CW283276 104_759.1	356	20.6	49.0	444	1	AI554349	AI554349 tq16c04.x
C 284	20.8	49.5	677	2	BG337825	BG337825 602435728	357	20.6	49.0	444	7	AW612046	AW612046 hg84c05.x
C 285	20.8	49.5	679	5	CK149312	CK149312 1742HFUN2	358	20.6	49.0	445	1	AI858307	AI858307 wl36b01.x
C 286	20.8	49.5	689	14	CT250268	CT250268 Sus scrof	359	20.6	49.0	446	1	AI005386	AI005386 ot98c02.x
C 287	20.8	49.5	694	4	BX923233	BX923233 BX923233	360	20.6	49.0	446	1	AA476600	AA476600 zx03d01.s
C 288	20.8	49.5	710	10	DV188880	DV188880 BCT032.G0	361	20.6	49.0	450	1	AI023935	AI023935 ow53g11.s
C 289	20.8	49.5	711	5	CF443933	CF443933 EST680278	362	20.6	49.0	450	7	BF445340	BF445340 7q81b06.x
C 290	20.8	49.5	737	10	DT182350	DT182350 JGI_ANN04	363	20.6	49.0	451	1	AI683228	AI683228 tx02b07.x
C 291	20.8	49.5	746	12	CG884531	CG884531 ZMMBB049	364	20.6	49.0	451	1	AA408540	AA408540 zv15f05.s
C 292	20.8	49.5	755	5	CF445140	CF445140 EST681485	365	20.6	49.0	452	7	AW192326	AW192326 xl14a03.x
C 293	20.8	49.5	755	12	CG185881	CG185881 PUFPLN16TB	366	20.6	49.0	453	7	AW466920	AW466920 ha04g08.x
294	20.8	49.5	756	12	CE441187	CE441187 tigr-gss-	367	20.6	49.0	454	1	AI922793	AI922793 wol14a05.x
C 295	20.8	49.5	769	12	CG826244	CG826244 ZMMBB017	C 368	20.6	49.0	455	2	BJ629249	BJ629249 BJ629249
C 296	20.8	49.5	770	8	CX136866	CX136866 AGENCOURT	369	20.6	49.0	457	1	AI799987	AI799987 wc46g03.x
C 297	20.8	49.5	786	5	CF447141	CF447141 EST683486	370	20.6	49.0	457	1	AI167426	AI167426 ox69h11.s
C 298	20.8	49.5	791	3	BUE94085	BUE94085 EST0122.A	371	20.6	49.0	458	1	AI613122	AI613122 ty71h03.x
C 299	20.8	49.5	797	4	CB312710	CB312710 AGENCOURT	372	20.6	49.0	458	1	AI690279	AI690279 tv33g01.x
C 300	20.8	49.5	801	10	DT562577	DT562577 EST107321	373	20.6	49.0	458	7	BE501326	BE501326 7a41a09.x
C 301	20.8	49.5	802	13	CZ253777	CZ253777 LOCy428TV	374	20.6	49.0	459	1	AA961556	AA961556 oq79f10.s
302	20.8	49.5	841	4	CA788238	CA788238 AGENCOURT	375	20.6	49.0	460	1	AI056434	AI056434 oy50c10.x
303	20.8	49.5	850	5	CK777835	CK777835 964980.MA	376	20.6	49.0	461	1	AI631810	AI631810 wa36d01.x
304	20.8	49.5	881	12	CG185883	CG185883 PUFPLN16TD	377	20.6	49.0	464	2	BG958837	BG958837 PM4-CT080
C 305	20.8	49.5	896	3	BE790548	BE790548 601582563	378	20.6	49.0	466	1	AI769143	AI769143 wg34d02.x
306	20.8	49.5	899	3	BQ928225	BQ928225 AGENCOURT	379	20.6	49.0	466	4	CB852976	CB852976 UI-CF-FNO
307	20.8	49.5	1031	3	BQ958966	BQ958966 AGENCOURT	380	20.6	49.0	467	7	AW502943	AW502943 UI-HF-BP0
308	20.8	49.5	1073	3	BUI89958	BUI89958 AGENCOURT	381	20.6	49.0	468	7	BE551997	BE551997 hy02d03.x
309	20.8	49.5	1204	12	CC260196	CC260196 CH261-531	C 382	20.6	49.0	468	10	H00496	H00496 y124c08.r1
C 310	20.6	49.0	206	7	AW811783	AW811783 RC3-ST016	383	20.6	49.0	470	1	AI608748	AI608748 tw90d02.x
C 311	20.6	49.0	280	7	BE768914	BE768914 PM1-FT002	384	20.6	49.0	470	7	AV755002	AV755002





C 531	20.6	49.0	898	3	BQ646878	BO646878 AGENCOURT	604	20.4	48.6	539	3	BM839320	BM839320 K-EST0116
C 532	20.6	49.0	899	3	BUI189396	BUI189396 AGENCOURT	605	20.4	48.6	539	3	DA796970	DA796970
C 533	20.6	49.0	900	3	DR6664118	DR6664118 ESTI05423	606	20.4	48.6	539	3	DA989776	DA989776
C 534	20.6	49.0	909	3	BUS08936	BUS08936 AGENCOURT	607	20.4	48.6	540	9	DA297709	DA297709
C 535	20.6	49.0	915	10	DR672682	DR672682 ESTI06279	608	20.4	48.6	540	9	DA773480	DA773480
C 536	20.6	49.0	919	7	BE885732	BE885732 601508944	609	20.4	48.6	540	9	DA867860	DA867860
C 537	20.6	49.0	922	10	DR671573	DR671573 ESTI06169	610	20.4	48.6	541	9	DA438005	DA438005
C 538	20.6	49.0	925	3	BUI165566	BUI165566 AGENCOURT	611	20.4	48.6	542	9	DB001254	DB001254
C 539	20.6	49.0	930	10	DR630257	DR630257 ESTI02038	612	20.4	48.6	542	9	DB122090	DB122090
C 540	20.6	49.0	939	10	DR665600	DR665600 ESTI05571	613	20.4	48.6	543	9	DA575647	DA575647
C 541	20.6	49.0	941	4	BX380892	BX380892 BX380892	614	20.4	48.6	543	9	DB258069	DB258069
C 542	20.6	49.0	948	3	BM802535	BM802535 AGENCOURT	615	20.4	48.6	544	9	DA276855	DA276855
C 543	20.6	49.0	954	1	AL550675	AL550675 AL550675	616	20.4	48.6	544	9	DA651258	DA651258
C 544	20.6	49.0	970	12	CC426405	CC426405 PUHFF57TD	617	20.4	48.6	546	9	DA758232	DA758232
C 545	20.6	49.0	978	10	DR753481	DR753481 CCAPAPA04	618	20.4	48.6	548	9	DA353888	DA353888
C 546	20.6	49.0	989	3	BM803520	BM803520 AGENCOURT	619	20.4	48.6	548	9	DA715772	DA715772
C 547	20.6	49.0	1027	3	BM903886	BM903886 AGENCOURT	620	20.4	48.6	549	9	DA120549	DA120549
C 548	20.6	49.0	1048	13	CL988825	CL988825 ZMMBH6000	621	20.4	48.6	549	9	DB129337	DB129337
C 549	20.6	49.0	1079	1	AL578222	AL578222 AL578222	622	20.4	48.6	550	10	DT673629	DT673629
C 550	20.6	49.0	1097	2	BG290255	BG290255 602384956	623	20.4	48.6	551	9	DA932100	DA932100
C 551	20.6	49.0	1139	6	CR609981	CR609981 full-leng	624	20.4	48.6	551	9	DA274458	DA274458
C 552	20.6	49.0	1689	14	DQ053049	DQ053049 Homo sapi	625	20.4	48.6	554	9	DB010322	DB010322
C 553	20.6	49.0	2048	6	CR602997	CR602997 full-leng	626	20.4	48.6	554	9	DB180987	DB180987
C 554	20.6	49.0	2752	6	BC014615	BC014615 Homo sapi	627	20.4	48.6	555	9	DA495232	DA495232
C 555	20.6	49.0	2785	6	CR857698	CR857698 Pongo pyg	628	20.4	48.6	556	9	DB104085	DB104085
C 556	20.6	49.0	5305	6	BC042194	BC042194 Homo sapi	629	20.4	48.6	557	9	DA159711	DA159711
C 557	20.4	48.6	196	13	CL655601	CL655601 PRI0124a_	630	20.4	48.6	557	9	DA295987	DA295987
C 558	20.4	48.6	239	7	BS584230	BS584230 BS584230	631	20.4	48.6	557	9	DA355653	DA355653
C 559	20.4	48.6	253	10	DT911113	DT911113 PCR830880	632	20.4	48.6	557	9	DA749462	DA749462
C 560	20.4	48.6	255	1	AJ769677	AJ769677 AJ769677	633	20.4	48.6	557	9	DB141283	DB141283
C 561	20.4	48.6	319	9	DA901604	DA901604 DA901604	634	20.4	48.6	558	9	DA724630	DA724630
C 562	20.4	48.6	342	4	C81714	C81714 C81714 Citr	635	20.4	48.6	558	9	DA739090	DA739090
C 563	20.4	48.6	371	5	CD700543	CD700543 ESTI17067	636	20.4	48.6	559	9	DA411398	DA411398
C 564	20.4	48.6	397	9	DA944834	DA944834 DA944834	637	20.4	48.6	559	9	DA589646	DA589646
C 565	20.4	48.6	404	10	DT053430	DT053430 COT_FQ_C0	638	20.4	48.6	559	11	AZ087885	AZ087885
C 566	20.4	48.6	429	9	DN619561	DN619561 UCRCS11_0	639	20.4	48.6	561	9	DA824486	DA824486
C 567	20.4	48.6	453	4	BY290880	BY290880 BY290880	640	20.4	48.6	561	9	DA875830	DA875830
C 568	20.4	48.6	458	10	DR648792	DR648792 ESTI03890	641	20.4	48.6	562	9	DA052390	DA052390
C 569	20.4	48.6	459	10	H01408	H01408 Y199b10_81	642	20.4	48.6	562	9	DA070759	DA070759
C 570	20.4	48.6	460	9	DA123254	DA123254 DA123254	643	20.4	48.6	562	9	DA410342	DA410342
C 571	20.4	48.6	460	9	DA184914	DA184914 DA184914	644	20.4	48.6	562	9	DA451541	DA451541
C 572	20.4	48.6	467	9	DB015241	DB015241 DB015241	645	20.4	48.6	562	13	CW084394	CW084394
C 573	20.4	48.6	472	10	DR623909	DR623909 ESTI01403	646	20.4	48.6	563	9	DA451844	DA451844
C 574	20.4	48.6	475	13	CZ362823	CZ362823 ZMMBF0141	647	20.4	48.6	563	9	DA451844	DA451844
C 575	20.4	48.6	484	2	BM153678	BM153678 TCBAPE12	648	20.4	48.6	564	9	DA450701	DA450701
C 576	20.4	48.6	486	9	DA197929	DA197929 DA197929	649	20.4	48.6	564	9	DA782343	DA782343
C 577	20.4	48.6	490	9	DA740499	DA740499 DA740499	650	20.4	48.6	565	9	DB128146	DB128146
C 578	20.4	48.6	490	9	DA741430	DA741430 DA741430	651	20.4	48.6	565	9	DA199685	DA199685
C 579	20.4	48.6	499	13	DU101816	DU101816 JBNY03900	652	20.4	48.6	566	9	DA452973	DA452973
C 580	20.4	48.6	500	9	DA741390	DA741390 DA741390	653	20.4	48.6	567	9	DA116268	DA116268
C 581	20.4	48.6	500	9	DA741557	DA741557 DA741557	654	20.4	48.6	567	9	DA208058	DA208058
C 582	20.4	48.6	503	9	DA889098	DA889098 DA889098	655	20.4	48.6	567	10	DR609565	DR609565
C 583	20.4	48.6	504	1	AJ497845	AJ497845 AJ497845	656	20.4	48.6	568	9	DA844736	DA844736
C 584	20.4	48.6	505	8	CN365809	CN365809 170006002	657	20.4	48.6	568	9	DA021703	DA021703
C 585	20.4	48.6	505	9	DA030327	DA030327 DA030327	658	20.4	48.6	569	9	DA397469	DA397469
C 586	20.4	48.6	506	9	DA554409	DA554409 DA554409	659	20.4	48.6	569	9	DA577068	DA577068
C 587	20.4	48.6	513	9	DA715433	DA715433 DA715433	660	20.4	48.6	569	9	DA922562	DA922562
C 588	20.4	48.6	516	10	DR623216	DR623216 ESTI01334	661	20.4	48.6	569	9	DA922562	DA922562
C 589	20.4	48.6	521	9	DA785287	DA785287 DA785287	662	20.4	48.6	569	9	DA386805	DA386805
C 590	20.4	48.6	522	9	DA739803	DA739803 DA739803	663	20.4	48.6	570	9	DA745119	DA745119
C 591	20.4	48.6	525	9	DA553880	DA553880 DA553880	664	20.4	48.6	570	9	DB051104	DB051104
C 592	20.4	48.6	525	9	DA578817	DA578817 DA578817	665	20.4	48.6	570	9	DB098316	DB098316
C 593	20.4	48.6	525	9	DA850979	DA850979 DA850979	666	20.4	48.6	571	9	DA297667	DA297667
C 594	20.4	48.6	525	9	DA923933	DA923933 DA923933	667	20.4	48.6	572	9	DB164397	DB164397
C 595	20.4	48.6	526	1	AL598146	AL598146 DKEFP313D	668	20.4	48.6	573	2	BI427518	BI427518
C 596	20.4	48.6	526	9	DA205495	DA205495 DA205495	669	20.4	48.6	573	9	DA391956	DA391956
C 597	20.4	48.6	527	10	DR625202	DR625202 ESTI01533	670	20.4	48.6	573	9	DA928209	DA928209
C 598	20.4	48.6	531	9	DA151114	DA151114 DA151114	671	20.4	48.6	574	9	DB151492	DB151492
C 599	20.4	48.6	532	9	DA745231	DA745231 DA745231	672	20.4	48.6	574	9	DA274851	DA274851
C 600	20.4	48.6	533	9	DA418152	DA418152 DA418152	673	20.4	48.6	574	9	DA564521	DA564521
C 601	20.4	48.6	536	8	CN365808	CN365808 170005326	674	20.4	48.6	574	9	DA690472	DA690472
C 602	20.4	48.6	536	9	DA621797	DA621797 DA621797	675	20.4	48.6	574	9	DB146555	DB146555
C 603	20.4	48.6	537	9	DA391932	DA391932 DA391932	676	20.4	48.6	574	9	DB160985	DB160985

677	20.4	48.6	575	9	DA752675	DA752675	C 750	20.4	48.6	661	8	CX053448
678	20.4	48.6	575	9	DB016042	DB016042	C 751	20.4	48.6	662	9	DN279867
679	20.4	48.6	575	9	DB198734	DB198734	752	20.4	48.6	662	13	CW176274
680	20.4	48.6	575	5	CD886611	G118.102M	753	20.4	48.6	663	8	CN365807
681	20.4	48.6	576	9	DA286467	DA286467	C 754	20.4	48.6	667	13	CZ648685
682	20.4	48.6	576	9	DA387063	DA387063	755	20.4	48.6	669	8	CV719946
683	20.4	48.6	576	9	DA924105	DA924105	756	20.4	48.6	672	8	CX079095
684	20.4	48.6	577	9	DB152847	DB152847	757	20.4	48.6	673	10	DR632354
685	20.4	48.6	579	9	DA296626	DA296626	758	20.4	48.6	674	10	DR6313673
686	20.4	48.6	580	9	DA204059	DA204059	759	20.4	48.6	676	2	B1768647
687	20.4	48.6	580	9	DA389723	DA389723	760	20.4	48.6	679	5	CF125451
688	20.4	48.6	580	9	DA389948	DA389948	761	20.4	48.6	679	8	CX075282
689	20.4	48.6	580	10	DR622902	EST101303	762	20.4	48.6	680	10	DR628789
690	20.4	48.6	580	10	DR632404	EST102253	763	20.4	48.6	681	5	CF125357
691	20.4	48.6	581	9	DA151584	DA151584	764	20.4	48.6	681	10	DR624097
692	20.4	48.6	581	9	DA436817	DA436817	765	20.4	48.6	682	10	DR623515
693	20.4	48.6	581	9	DB259366	DB259366	766	20.4	48.6	682	10	DR623515.C
694	20.4	48.6	582	9	DA301416	DA301416	C 767	20.4	48.6	683	8	CX075281
695	20.4	48.6	582	9	DA863767	DA863767	768	20.4	48.6	684	10	DR625159
696	20.4	48.6	585	9	DA292745	DA292745	769	20.4	48.6	686	14	AG121405
697	20.4	48.6	585	9	DA345706	DA345706	770	20.4	48.6	694	13	CW309218
698	20.4	48.6	586	2	BG706909	602672082	771	20.4	48.6	698	14	EX209168
699	20.4	48.6	586	9	CX554931	Yd893f12.	772	20.4	48.6	698	14	EX209168
700	20.4	48.6	586	9	DA988895	DA988895	773	20.4	48.6	701	7	BE731836
701	20.4	48.6	587	9	DA116506	DA116506	C 774	20.4	48.6	710	12	CC717906
702	20.4	48.6	587	9	DA128497	DA128497	775	20.4	48.6	711	10	DR630145
703	20.4	48.6	587	9	DA281071	DA281071	776	20.4	48.6	711	11	BZ038465
704	20.4	48.6	589	9	DA497274	DA497274	777	20.4	48.6	711	13	CW309219
705	20.4	48.6	591	9	DA771098	DA771098	778	20.4	48.6	712	10	DR624004
706	20.4	48.6	591	9	DB163011	DB163011	779	20.4	48.6	712	12	CL193312
707	20.4	48.6	592	5	CF766301	CES000248	780	20.4	48.6	715	9	DN280932
708	20.4	48.6	594	9	DA676094	DA676094	781	20.4	48.6	720	10	DR622588
709	20.4	48.6	594	9	DB031221	DB031221	782	20.4	48.6	724	10	DR624030
710	20.4	48.6	594	9	DB098497	DB098497	783	20.4	48.6	724	10	DR624158
711	20.4	48.6	595	9	DA204383	DA204383	784	20.4	48.6	728	1	AU130485
C 712	20.4	48.6	595	10	DR607747	EST937875	785	20.4	48.6	731	12	CL155428
713	20.4	48.6	599	9	DA408949	DA408949	786	20.4	48.6	735	4	CA766731
714	20.4	48.6	599	9	DA536446	DA536446	787	20.4	48.6	738	7	BE733723
C 715	20.4	48.6	599	11	AZ602905	AZ602905	788	20.4	48.6	740	10	DR631677
716	20.4	48.6	600	9	DB164566	DB164566	789	20.4	48.6	745	10	DR631677
C 717	20.4	48.6	601	4	CA144169	CA144169	790	20.4	48.6	750	8	CN365811
718	20.4	48.6	604	9	DA111137	DA111137	C 791	20.4	48.6	751	3	BU444563
719	20.4	48.6	605	11	AZ736059	RFC1-24-1	792	20.4	48.6	758	13	CL486136
C 720	20.4	48.6	609	9	CX638278	UCRPT02.3	793	20.4	48.6	761	2	BG753122
721	20.4	48.6	609	10	DR624446	EST101457	794	20.4	48.6	766	10	DR629331
C 722	20.4	48.6	610	10	DR611911	DR611911	795	20.4	48.6	770	12	CC346619
723	20.4	48.6	613	9	DB031613	DB031613	796	20.4	48.6	775	5	CJ456626
724	20.4	48.6	615	9	DA530469	DA530469	797	20.4	48.6	776	10	DR622257
725	20.4	48.6	620	7	AW239216	AW239216	798	20.4	48.6	779	10	DR622257
726	20.4	48.6	626	8	CN365810	CN365810	799	20.4	48.6	781	10	DR614730
727	20.4	48.6	627	2	BG722171	BG722171	C 800	20.4	48.6	784	4	CB570527
728	20.4	48.6	629	10	DR631628	DR631628	C 801	20.4	48.6	791	10	DR626277
729	20.4	48.6	630	9	DA373137	DA373137	802	20.4	48.6	806	2	B1548671
730	20.4	48.6	632	10	DR623304	DR623304	C 803	20.4	48.6	810	13	CL399468
731	20.4	48.6	632	10	DR630991	DR630991	804	20.4	48.6	817	8	CO403906
732	20.4	48.6	632	10	DR632360	DR632360	C 805	20.4	48.6	832	10	DT085020
733	20.4	48.6	633	8	CN365813	CN365813	C 806	20.4	48.6	833	10	DT085020
C 734	20.4	48.6	635	2	CV710305	CV710305	C 807	20.4	48.6	834	9	DN288236
735	20.4	48.6	642	2	BG715825	BG715825	C 808	20.4	48.6	835	13	DU151339
736	20.4	48.6	642	10	DR630366	DR630366	C 809	20.4	48.6	838	9	CV561813
737	20.4	48.6	643	4	BX498357	BX498357	810	20.4	48.6	838	9	DA197292
C 738	20.4	48.6	650	8	CV719945	CV719945	811	20.4	48.6	839	10	DR628297
739	20.4	48.6	652	2	BG333305	BG333305	812	20.4	48.6	841	12	CC628320
740	20.4	48.6	652	8	CN365800	CN365800	C 813	20.4	48.6	842	14	CT471732
741	20.4	48.6	653	5	CK937461	CK937461	814	20.4	48.6	849	10	DR626504
742	20.4	48.6	654	8	CV710306	CV710306	815	20.4	48.6	852	10	DR630381
743	20.4	48.6	657	8	CX053449	CX053449	816	20.4	48.6	854	3	BQ213153
744	20.4	48.6	657	13	CW081126	CW081126	817	20.4	48.6	858	3	BU526982
745	20.4	48.6	658	8	CN915192	CN915192	818	20.4	48.6	859	9	DN511381
- 746	20.4	48.6	658	10	DR613425	DR613425	819	20.4	48.6	861	12	CC717912
C 747	20.4	48.6	659	4	CB168018	CB168018	820	20.4	48.6	872	10	DR681394
748	20.4	48.6	660	8	CT000578	CT000578	821	20.4	48.6	886	12	BZ792501
749	20.4	48.6	661	4	CA858155	EST635410	822	20.4	48.6	889	3	BU528156

CX053448 UCRCS09.6  
DN279867 1160928\_M  
CW176274 104\_589\_1  
CN365807 170004240  
CZ648685 OM\_Ba019  
CV719946 UCRCS08.0  
CX079095 yd807e1t  
DR632354 EST102248  
DR6313673 EST100380  
B1768647 603056989  
CF125451 UI\_HF-EL0  
CX075282 UCRCS08.4  
DR628789 EST101891  
CF125357 UI\_HF-EL0  
DR624097 EST101422  
DR623515 EST101364  
DX871575 LB02515.C  
CX075281 UCRCS08.4  
DR625159 EST101528  
AG121405 Pan trogl  
CW309218 104\_798\_1  
EX209168 Danilo rer  
DR624004 EST101413  
BE731836 601565890  
CC717906 OGLAI45TH  
DR630145 EST102027  
BZ038465 1j042a11.  
CW309219 104\_798\_1  
DV876369 LB0255.CR  
CL193312 104\_416\_1  
DN280932 1162080\_M  
DR622588 EST101271  
DR624030 EST101415  
DR624158 EST101428  
AU130485 AU130485  
CL155428 104\_341\_1  
CA766731 AF53-Rof\_1  
BE733723 601568240  
DR631677 EST102180  
DT846508 LB00473.C  
CN365811 170005315  
BU444563 603763630  
CL486136 SATL\_428  
BG753122 602732394\_-  
DR629331 EST101945  
CC346619 OGNBT14TV  
CJ456626 CJ456626  
DR622257 EST101238  
DV907757 LB0283.CR  
DR614730 EST100485  
CB570527 AGENCOURT  
DR626277 EST101640  
B1548671 603196841  
CL399468 ZMMBB040  
CO403906 AGENCOURT  
DT085020 JGI\_A1ZM4  
DV760917 FchrSQ01  
DN288236 1242726\_M  
DU151339 OG\_ABA003  
CV561813 UI-N-HL0-  
DA197292 DA197292  
DR628297 EST101842  
CC628320 OGLAK44TV  
CT471732 Sus scrofa  
DR626504 EST101663  
DR630381 EST102050  
BQ213153 AGENCOURT  
BU526982 AGENCOURT  
DN511381 1242462\_M  
CC717912 OGLAI45TV  
DV881394 LB02618.C  
BZ792501 PUFGP79TD  
BU528156 AGENCOURT

823	20.4	48.6	890	3	B0924082	B0924082 AGENCOURT	896	20.2	48.1	607	12	CC343310	CC343310 OGSARH5TV
824	20.4	48.6	894	3	B0953656	B0953656 AGENCOURT	C 897	20.2	48.1	609	13	CZ660671	CZ660671 OM_BA021
825	20.4	48.6	906	8	CR990019	CR990019 CRN258-G0	C 898	20.2	48.1	610	13	DU319079	DU319079 109844800
826	20.4	48.6	907	10	DV009608	DV009608 CRN258-G0	C 899	20.2	48.1	611	11	AQ983921	AQ983921 RPCI-23-3
827	20.4	48.6	909	3	B0931683	B0931683 AGENCOURT	900	20.2	48.1	616	7	AW226590	AW226590 up0907.x
828	20.4	48.6	911	3	B0932003	B0932003 AGENCOURT	901	20.2	48.1	630	11	AZ608248	AZ608248 1M0432010
829	20.4	48.6	912	3	B0931341	B0931341 AGENCOURT	C 902	20.2	48.1	638	11	BH025261	BH025261 RPCI-24-3
830	20.4	48.6	913	13	DU161500	DU161500 OG_ABA004	C 903	20.2	48.1	649	13	CM151384	CM151384 104_552_1
831	20.4	48.6	914	3	B0927085	B0927085 AGENCOURT	C 904	20.2	48.1	650	13	CW275554	CW275554 104_748_1
832	20.4	48.6	922	3	B0844726	B0844726 AGENCOURT	905	20.2	48.1	651	11	AZ407349	AZ407349 1M0178D06
833	20.4	48.6	924	3	B0926599	B0926599 AGENCOURT	C 906	20.2	48.1	662	14	LBFA062C03	LBFA062C03
834	20.4	48.6	929	4	BA460464	BA460464 BA460464	C 907	20.2	48.1	667	14	LBFA062C03	LBFA062C03
835	20.4	48.6	934	1	AL551372	AL551372	C 908	20.2	48.1	674	12	CC313716	CC313716 TAM32-121
836	20.4	48.6	938	3	B0926908	B0926908 AGENCOURT	C 909	20.2	48.1	687	12	CG401453	CG401453 ZWMBB024
837	20.4	48.6	938	12	CG458887	CG458887 PUJUV31TD	910	20.2	48.1	708	8	CA458559	CA458559 UI-M-HNO-
838	20.4	48.6	944	12	CG458850	CG458850 PUJUV31TB	911	20.2	48.1	710	4	CB248238	CB248238 UI-M-F10-
839	20.4	48.6	947	3	B0931097	B0931097 AGENCOURT	912	20.2	48.1	710	5	CF737766	CF737766 UI-M-HDO-
840	20.4	48.6	961	5	CD515346	CD515346 AGENCOURT	913	20.2	48.1	719	10	DT894056	DT894056 1476635 M
841	20.4	48.6	962	3	B0947263	B0947263 AGENCOURT	914	20.2	48.1	720	11	BH657012	BH657012 BOMHP54TF
842	20.4	48.6	970	1	AL528843	AL528843	C 915	20.2	48.1	721	12	CG066252	CG066252 PUKDT40TD
843	20.4	48.6	987	3	B0952082	B0952082 AGENCOURT	916	20.2	48.1	732	9	CX383744	CX383744 JGI_XZT63
844	20.4	48.6	988	1	AL583301	AL583301	917	20.2	48.1	738	5	CF728738	CF728738 UI-M-HDO-
845	20.4	48.6	999	4	CA493956	CA493956 AGENCOURT	918	20.2	48.1	739	8	CV559161	CV559161 UI-M-H20-
846	20.4	48.6	1018	12	CG195859	CG195859 PUJDO82TB	919	20.2	48.1	740	14	CNS010MC	AL153700 Anopheles
847	20.4	48.6	1058	3	B0923321	B0923321 AGENCOURT	920	20.2	48.1	744	5	CK635148	CK635148 UI-M-HNO-
848	20.4	48.6	1130	2	B0467075	B0467075 AGENCOURT	921	20.2	48.1	744	8	CNS32742	CNS32742 UI-M-HDO-
849	20.4	48.6	1149	3	B0277813	B0277813 AGENCOURT	922	20.2	48.1	746	13	CL865019	CL865019 t1k17b6.f
850	20.4	48.6	1156	14	CNS06PRC	AL407632 Pan trogl	923	20.2	48.1	746	13	CL838742	CL838742 OR_CBA006
851	20.4	48.6	1158	3	B0563167	B0563167 AGENCOURT	C 924	20.2	48.1	757	12	CG602501	CG602501 OGULU49TV
852	20.4	48.6	1260	14	AY407632	AY407632 Pan trogl	925	20.2	48.1	765	8	CO402993	CO402993 AGENCOURT
853	20.4	48.6	1323	14	AY407631	AY407631 Homo sapi	926	20.2	48.1	766	8	CN461648	CN461648 UI-M-HNO-
854	20.4	48.6	1436	12	CC202002	CC202002 CH261-311	927	20.2	48.1	772	13	CL747392	CL747392 OR_BA011
855	20.2	48.1	161	1	AL702040	AL702040 tq20f05.x	C 928	20.2	48.1	782	11	BZ083809	BZ083809 11a18d04.
856	20.2	48.1	181	5	CF561492	CF561492 1115065G1	929	20.2	48.1	787	13	CL854824	CL854824 UI-M-G10-
857	20.2	48.1	202	2	BM288720	BM288720 530102 MA	930	20.2	48.1	792	4	CB519934	CB519934 UI-M-G10-
858	20.2	48.1	290	1	AA017118	AA017118 z33g09.r	C 931	20.2	48.1	792	14	CR119685	CR119685 Reverse S
859	20.2	48.1	290	8	CO436193	CO436193 d6PCR16.1	C 932	20.2	48.1	806	10	DW252048	DW252048 UI-S-CB0-
860	20.2	48.1	317	4	BW888782	BW888782	C 933	20.2	48.1	848	10	DT827919	DT827919 LB00254.C
861	20.2	48.1	318	7	BF652500	BF652500 27080 MA	C 934	20.2	48.1	854	2	B1077759	B1077759 602871519
862	20.2	48.1	328	5	CJ272698	CJ272698 CJ272698	935	20.2	48.1	855	2	BM451888	BM451888 AGENCOURT
863	20.2	48.1	356	11	AQ934814	AQ934814 RPCI-23-2	C 941	20.2	48.1	864	11	AQ576152	AQ576152 nxbx0088N
864	20.2	48.1	359	7	BE668226	BE668226 156816 MA	C 942	20.2	48.1	864	14	DU910609	DU910609 332362 To
865	20.2	48.1	387	1	AI666501	AI666501 mu21a07.x	943	20.2	48.1	866	8	CR995417	CR995417 CR995417
866	20.2	48.1	394	5	CJ324534	CJ324534 CJ324534	944	20.2	48.1	878	2	BG027188	BG027188 602295874
867	20.2	48.1	396	11	AQ671332	AQ671332 HS_5452.A	945	20.2	48.1	879	5	CF812070	CF812070 EST689452
868	20.2	48.1	401	2	BF957130	BF957130 QV4-NM114	C 946	20.2	48.1	883	5	CK414374	CK414374 AUF_IpGill
869	20.2	48.1	407	10	H77921	H77921 ysl0b07.s1	947	20.2	48.1	883	8	CNS02349	CNS02349 AGENCOURT
870	20.2	48.1	409	1	AA274204	AA274204 vb92b07.r	948	20.2	48.1	899	2	BX109793	BX109793 602280945
871	20.2	48.1	412	11	AZ477800	AZ477800 1M0297D10	949	20.2	48.1	912	4	BX366351	BX366351 BX366351
872	20.2	48.1	415	12	BF811492	BF811492 RC6-C1000	C 950	20.2	48.1	929	5	CK425244	CK425244 AUF_IpSto
873	20.2	48.1	418	12	CL174817	CL174817 104_379.1	C 951	20.2	48.1	932	3	BU854542	BU854542 AGENCOURT
874	20.2	48.1	428	7	BF652493	BF652493 276066 MA	C 952	20.2	48.1	936	14	DU888204	DU888204 389797 To
875	20.2	48.1	430	2	BF810195	BF810195 RC6-C1000	C 953	20.2	48.1	990	13	DU010982	DU010982 229310 To
876	20.2	48.1	439	1	AA137870	AA137870 mq29e01.r	954	20.2	48.1	991	13	CL494234	CL494234 SAIL_590
877	20.2	48.1	442	7	AW494740	AW494740 UI-M-BH3-	C 955	20.2	48.1	999	12	CG103422	CG103422 PUJBV65TB
878	20.2	48.1	448	1	AI800142	AI800142 tr22h01.x	C 956	20.2	48.1	1012	7	BF317070	BF317070 601903468
879	20.2	48.1	458	1	AL921109	AL921109	C 957	20.2	48.1	1031	4	BY709578	BY709578 BY709578
880	20.2	48.1	465	3	BU817657	BU817657 UAL9CEP09	C 958	20.2	48.1	1031	4	BY709578	BY709578 BY709578
881	20.2	48.1	475	1	AL828284	AL828284	C 959	20.2	48.1	1031	4	BY709578	BY709578 BY709578
882	20.2	48.1	480	1	AA981888	AA981888 ua34a08.r	C 960	20.2	48.1	1031	4	BY709578	BY709578 BY709578
883	20.2	48.1	482	11	AQ880972	AQ880972 HS_5106.B	C 961	20.2	48.1	1031	4	BY709578	BY709578 BY709578
884	20.2	48.1	492	1	AI510554	AI510554 mq29e01.y	C 962	20.2	48.1	1031	4	BY709578	BY709578 BY709578
885	20.2	48.1	497	7	AW993833	AW993833 RC3-BN003	C 963	20.2	48.1	1031	4	BY709578	BY709578 BY709578
886	20.2	48.1	507	1	AI639212	AI639212 rx009208	C 964	20.2	48.1	1031	4	BY709578	BY709578 BY709578
887	20.2	48.1	535	11	AQ507786	AQ507786 RPCI-11-3	C 965	20.2	48.1	1031	4	BY709578	BY709578 BY709578
888	20.2	48.1	537	14	BX131335	BX131335 Danio rer	C 966	20.2	48.1	1031	4	BY709578	BY709578 BY709578
889	20.2	48.1	546	11	AZ616384	AZ616384 1M0446K08	C 967	20.2	48.1	1031	4	BY709578	BY709578 BY709578
890	20.2	48.1	551	11	BH747824	BH747824 SALK_0352	C 968	20.2	48.1	1031	4	BY709578	BY709578 BY709578
891	20.2	48.1	565	10	DT903345	DT903345 S29-21519	C 969	20.2	48.1	1031	4	BY709578	BY709578 BY709578
892	20.2	48.1	584	4	CB245904	CB245904 UI-M-FOO-	C 970	20.2	48.1	1031	4	BY709578	BY709578 BY709578
893	20.2	48.1	596	11	AZ458480	AZ458480 1M0262B11	C 971	20.2	48.1	1031	4	BY709578	BY709578 BY709578
894	20.2	48.1	596	13	CZ630362	CZ630362 OM_Ba017	C 972	20.2	48.1	1031	4	BY709578	BY709578 BY709578
895	20.2	48.1	604	11	AQ482023	AQ482023 RPCI-11-2	C 973	20.2	48.1	1031	4	BY709578	BY709578 BY709578

```

969 20.2 48.1 5433 14 AY406469
970 20.2 48.1 6003 6 AK029889 Mus muscu
971 20.2 48.1 8207 6 AK147591 Mus muscu
972 20 47.6 126 1 AA750976
973 20 47.6 162 14 CR254107 Forward s
974 20 47.6 187 1 AA749619 ISAJ0092
975 20 47.6 235 2 BJ663789 BJ663789
976 20 47.6 266 7 BB049175 BB049175
977 20 47.6 269 3 BU095930 tca-415 t
978 20 47.6 274 8 CO362116 RTK1.1 H0
979 20 47.6 275 7 BB441961 BB441961
980 20 47.6 277 7 BB197214 BB197214
981 20 47.6 280 7 BB439695 BB439695
982 20 47.6 301 7 BB125964 BB125964
983 20 47.6 303 7 BB189370 BB189370
984 20 47.6 311 2 BG931284 BG931284
985 20 47.6 320 7 BB209177 BB209177
986 20 47.6 321 7 BB460781 BB460781
987 20 47.6 324 2 EG931869 EG931869
988 20 47.6 326 7 BB321388 BB321388
989 20 47.6 328 7 BB557115 BB557115
990 20 47.6 331 2 BG930982 BG930982
991 20 47.6 335 1 A1707155 T6796 MVA
992 20 47.6 349 7 AV754927 AV754927
993 20 47.6 388 4 BY635699 BY635699
994 20 47.6 400 9 DN802613 G.hir-15-
995 20 47.6 445 4 CA967782 CcLL03a12
996 20 47.6 457 5 CF140081 UI-HF-CB0
997 20 47.6 469 9 DN802398 G.hir-15-
998 20 47.6 480 10 DR455490 CM034G09
999 20 47.6 496 9 DR071068 RTDK1.17-
1000 20 47.6 511 14 CT142141 Sus scrof

ALIGNMENTS

RESULT 1
BI063311 158 bp mRNA linear EST 15-JUN-2001
LOCUS IL3-UT0117-300301-538-E01 UT0117 Homo sapiens cDNA, mRNA sequence.
DEFINITION BI063311
ACCESSION BI063311.1 GI:14470838
VERSION EST.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
1 (bases 1 to 158)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongenseel,C.V.,
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
10737800
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?t1=IL3&t2=IL3-UT0117-
300301-538-E01&t3=2001-03-30&t4=1)
Seq primer: puc 18 forward

TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
PUBMED 10737800
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?t1=IL3&t2=IL3-UT0117-
300301-538-E01&t3=2001-03-30&t4=1)
Seq primer: puc 18 forward

FEATURES
source
AY406469 Mus muscu
AK029889 Mus muscu
AK147591 Mus muscu
AA750976 ISKP0130
CR254107 Forward s
AA749619 ISAJ0092
BJ663789 BJ663789
BB049175 BB049175
BU095930 tca-415 t
CO362116 RTK1.1 H0
BB441961 BB441961
BB197214 BB197214
BB439695 BB439695
BB125964 BB125964
BB189370 BB189370
BG931284 BG931284
BB209177 BB209177
BB460781 BB460781
EG931869 EG931869
BB321388 BB321388
BB557115 BB557115
BG930982 BG930982
A1707155 T6796 MVA
AV754927 AV754927
BY635699 BY635699
DN802613 G.hir-15-
CA967782 CcLL03a12
CF140081 UI-HF-CB0
DN802398 G.hir-15-
DR455490 CM034G09
DR071068 RTDK1.17-
CT142141 Sus scrof

High quality sequence stop: 158.
Location/Qualifiers
1. .158
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="UT0117"
/note="Organ: uterus tumor; Vector: puc18; Site 1: SmaI;
Site 2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

ORIGIN
Query Match 100.0%; Score 42; DB 2; Length 158;
Best Local Similarity 100.0%; Pred. No. 1e-06;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACGCTTGCGCGAGCCCTTACAAAGACTATGAAGTAAAG 42
|||||
DB 100 GCCACGCTTGCGCGAGCCCTTACAAAGACTATGAAGTAAAG 141
|||||

RESULT 2
BE818833/c
LOCUS CM2-BN0302-050700-256-c01 BN0302 Homo sapiens cDNA, mRNA sequence.
DEFINITION BE818833
ACCESSION BE818833.1 GI:10251067
VERSION EST.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
1 (bases 1 to 307)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongenseel,C.V.,
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
10737800
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?t1=CM2-BN0302-050
700-256-c01&t3=2000-07-05&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 25
High quality sequence start: 307.

FEATURES
source
1. .307
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="BN0302"
/note="Organ: breast normal; Vector: puc18; Site 1: SmaI;
Site 2: SmaI; A mini-library was made by cloning products

```

derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

## ORIGIN

Query Match 100.0%; Score 42; DB 7; Length 307;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-06;  
 Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
 |||  
 Db 242 GCCCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 201

## RESULT 3

BE818817/c  
 LOCUS CM2-BN0302-050700-256-b07 BN0302 Homo sapiens cDNA, mRNA linear EST 21-SEP-2000  
 DEFINITION  
 ACCESSION BE818817  
 VERSION BE818817.1 GI:10251051  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS  
 1 (bases 1 to 341)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 341)  
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.  
 Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

TITLE  
 JOURNAL  
 PUBMED 10737800  
 COMMENT  
 Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2=CM2-BN0302-050700-256-b07&t3=2000-07-05&t4=1)  
 Seq primer: puc 18 forward  
 High quality sequence start: 10  
 High quality sequence stop: 341.

## FEATURES

source  
 1..341  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /dev\_stage="Adult"  
 /clone\_lib="BN0302"  
 /note="Organ: breast normal; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

## ORIGIN

Query Match 100.0%; Score 42; DB 7; Length 341;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-06;  
 Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
 |||  
 Db 253 GCCCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 212

## RESULT 4

BE818835/c  
 LOCUS CM2-BN0302-050700-256-cl1 BN0302 Homo sapiens cDNA, mRNA linear EST 21-SEP-2000  
 DEFINITION  
 ACCESSION BE818835  
 VERSION BE818835.1 GI:10251069  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS  
 1 (bases 1 to 384)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 384)  
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.  
 Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

TITLE  
 JOURNAL  
 PUBMED 10737800  
 COMMENT  
 Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2=CM2-BN0302-050700-256-cl1&t3=2000-07-05&t4=1)  
 Seq primer: puc 18 forward  
 High quality sequence start: 5  
 High quality sequence stop: 384.

## FEATURES

source  
 1..384  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /dev\_stage="Adult"  
 /clone\_lib="BN0302"  
 /note="Organ: breast normal; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

## ORIGIN

Query Match 100.0%; Score 42; DB 7; Length 384;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-06;  
 Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
 |||  
 Db 257 GCCCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 216

## RESULT 5

BE818828/c  
 LOCUS CM2-BN0302-050700-256-a05 BN0302 Homo sapiens cDNA, mRNA linear EST 21-SEP-2000  
 DEFINITION

```

ACCESSION  BE818828
VERSION     BE818828.1  GI:10251062
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homiidae; Homo.
REFERENCE  1 (bases 1 to 396)
AUTHORS   Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
            O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
PUBMED   10737800
COMMENT   Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2=CM2-BN0302-050
            700-256-f02&t3=2000-07-05&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 11
            High quality sequence stop: 396.
FEATURES   Location/Qualifiers
            source
            1..396
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /dev_stage="Adult"
                /clone_lib="BN0302"
                /note="Organ: breast normal; Vector: puc18; Site 1: SmaI;
                Site 2: SmaI; A mini-library was made by cloning products
                derived from ORESTES PCR (U.S. Letters Patent application
                No. 196,716 - Ludwig Institute for Cancer Research)
                profiles into the pUC 18 vector. Reverse transcription of
                tissue mRNA and cDNA amplification were performed under
                low stringency conditions."
ORIGIN
Query Match 100.0%; Score 42; DB 7; Length 396;
Best Local Similarity 100.0%; Pred. No. 1.2e-06;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAGCTAAAG 42
    |||||||||||||||||||||||||||||||||||||||
Db 248 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAGCTAAAG 207

RESULT 6
BE818840/c
LOCUS      CM2-BN0302-050700-256-f02 BN0302 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE818840
VERSION    BE818840.1  GI:10251074
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homiidae; Homo.
REFERENCE  1 (bases 1 to 404)
AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
            O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
PUBMED   10737800
COMMENT   Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2=CM2-BN0302-050
            700-256-f02&t3=2000-07-05&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 11
            High quality sequence stop: 396.
FEATURES   Location/Qualifiers
            source
            1..404
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /dev_stage="Adult"
                /clone_lib="BN0302"
                /note="Organ: breast normal; Vector: puc18; Site 1: SmaI;
                Site 2: SmaI; A mini-library was made by cloning products
                derived from ORESTES PCR (U.S. Letters Patent application
                No. 196,716 - Ludwig Institute for Cancer Research)
                profiles into the pUC 18 vector. Reverse transcription of
                tissue mRNA and cDNA amplification were performed under
                low stringency conditions."
ORIGIN
Query Match 100.0%; Score 42; DB 7; Length 404;
Best Local Similarity 100.0%; Pred. No. 1.2e-06;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAGCTAAAG 42
    |||||||||||||||||||||||||||||||||||||||
Db 244 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAGCTAAAG 203

RESULT 7
BE818805/c
LOCUS      CM2-BN0302-040700-253-b10 BN0302 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE818805
VERSION    BE818805.1  GI:10251039
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homiidae; Homo.
REFERENCE  1 (bases 1 to 424)
AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
            O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
PUBMED   10737800
COMMENT   Contact: Simpson A.J.G.

```

```

Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
10737800
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2=CM2-BN0302-050
700-256-f02&t3=2000-07-05&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 27
High quality sequence stop: 404.
FEATURES   Location/Qualifiers
            source
            1..404
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /dev_stage="Adult"
                /clone_lib="BN0302"
                /note="Organ: breast normal; Vector: puc18; Site 1: SmaI;
                Site 2: SmaI; A mini-library was made by cloning products
                derived from ORESTES PCR (U.S. Letters Patent application
                No. 196,716 - Ludwig Institute for Cancer Research)
                profiles into the pUC 18 vector. Reverse transcription of
                tissue mRNA and cDNA amplification were performed under
                low stringency conditions."
ORIGIN
Query Match 100.0%; Score 42; DB 7; Length 404;
Best Local Similarity 100.0%; Pred. No. 1.2e-06;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAGCTAAAG 42
    |||||||||||||||||||||||||||||||||||||||
Db 244 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAGCTAAAG 203

RESULT 7
BE818805/c
LOCUS      CM2-BN0302-040700-253-b10 BN0302 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE818805
VERSION    BE818805.1  GI:10251039
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homiidae; Homo.
REFERENCE  1 (bases 1 to 424)
AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
            O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
PUBMED   10737800
COMMENT   Contact: Simpson A.J.G.

```

Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2-cm2-BN0302-040  
700-253-b10&t3=2000-07-04&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 4  
High quality sequence stop: 424.  
Location/Qualifiers

#### FEATURES

source

1...424  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/dev\_stage="Adult"  
/clone\_lib="BN0302"  
/note="Organ: breast normal; Vector: puc18; Site 1: SmaI;  
Site 2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

#### ORIGIN

Query Match 100.0%; Score 42; DB 7; Length 424;  
Best Local Similarity 100.0%; Pred. No. 1.3e-06;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GCCCAGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
Db 250 GCCCAGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 209

#### RESULT 8

CD672930 453 bp mRNA linear EST 24-JUN-2003  
LOCUS fg18a05.y2 Human Iris cDNA (Normalized): fg Homo sapiens cDNA clone  
fg18a05 5', mRNA sequence.  
ACCESSION CD672930  
VERSION CD672930.1 GI:32174661  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
REFERENCE 1 (bases 1 to 453)  
AUTHORS Wistow,G., Bernstein,S.L., Ray,S., Wyatt,M.K., Behal,A.,  
Touchman,J.W., Bouffard,G., Smith,D. and Peterson,K.  
TITLE Expressed sequence tag analysis of adult human iris for the NEIBank  
Project: steroid-response factors and similarities with retinal  
pigment epithelium  
JOURNAL Mol. Vis. 8 (4), 185-195 (2002)  
PUBMED 12107412  
COMMENT Contact: Wistow G  
Section on Molecular Structure and Function  
National Eye Institute  
6/331, NIH, Bethesda, MD 20892-2740, USA  
Tel: 301 402 3452  
Fax: 301 496 0078  
Email: graeme@helix.nih.gov  
Plate: 18 row: a column: 05  
Seq primer: M13RP1 reverse primer (ABI).  
Location/Qualifiers

#### FEATURES

source

1..453  
/organism="Homo sapiens"  
/mol\_type="mRNA"

/db\_xref="taxon:9606"  
/clone\_lib="fg18a05"  
/tissue\_type="Iris"  
/dev\_stage="Adult"  
/lab\_host="EMDH10B"  
/clone\_lib="Human Iris cDNA (Normalized): fg"  
/note="Organ: Eye; Vector: pCMVSPORT6; A human iris  
library (ix) was normalized by self-subtraction. One  
portion of double stranded plasmid DNA representing the  
library was linearized by NotI. This NotI digested library  
was used as a template for biotinylated RNA synthesis  
using SP6 RNA polymerase. Another portion of the double  
stranded plasmid library was converted to single-stranded  
circles in vitro using Gene II and Exonuclease III (Life  
Technologies). Single-stranded DNA (1 mg) was hybridized  
(Cot 500) with 41 mg of Bio-RNA and vector blocking  
oligonucleotides. The hybridized Bio-RNA/ss-circles were  
removed by streptavidin:phenol extraction. EST analysis  
was performed on the library at the NIH Intramural  
Sequencing Center (NISC)."

#### ORIGIN

Query Match 100.0%; Score 42; DB 5; Length 453;  
Best Local Similarity 100.0%; Pred. No. 1.3e-06;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GCCCAGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
Db 56 GCCCAGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 97

#### RESULT 9

BE818871/c 458 bp mRNA linear EST 21-SEP-2000  
LOCUS BE818871  
DEFINITION CM2-BN0302-100700-259-f01 BN0302 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BE818871  
VERSION BE818871.1 GI:10251105  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
REFERENCE 1 (bases 1 to 458)  
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,  
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,  
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.  
TITLE Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
PUBMED 10737800  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2-cm2-BN0302-100  
700-259-f01&t3=2000-07-10&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 12  
High quality sequence stop: 458.  
Location/Qualifiers

#### FEATURES

source

1..458  
/organism="Homo sapiens"  
/mol\_type="mRNA"

```

/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="BN0302"
/notes="Organ: breast normal; Vector: puc18; Site 1: SmaI;
Site 2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

```

## ORIGIN

```

Query Match      100.0%; Score 42; DB 7; Length 458;
Best Local Similarity 100.0%; Pred. No. 1.3e-06;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAGTAAAG 42
    |||||||
Db 264 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAGTAAAG 223

```

## RESULT 10

```

BE818859/c
LOCUS              466 bp      mRNA      linear      EST 21-SEP-2000
DEFINITION      CM2-BN0302-100700-259-c02 BN0302 Homo sapiens cDNA, mRNA sequence.
ACCESSION      BE818859
VERSION      BE818859.1 GI:10251093
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens

```

```

REFERENCE
AUTHORS      Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V.,
O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.

```

```

TITLE      Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags

```

```

JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
PUBMED      10737800
COMMENT      Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=ct2=CM2-BN0302-100
700-259-c02&t3=2000-07-10&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 466.
Location/Qualifiers
1..466
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="BN0302"
/notes="Organ: breast normal; Vector: puc18; Site 1: SmaI;
Site 2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

```

```

FEATURES
source

```

```

Query Match      100.0%; Score 42; DB 7; Length 458;
Best Local Similarity 100.0%; Pred. No. 1.3e-06;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

## FEATURES source

```

Qy 1 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAGTAAAG 42
    |||||||
Db 246 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAGTAAAG 205

```

## ORIGIN

```

Query Match      100.0%; Score 42; DB 7; Length 466;
Best Local Similarity 100.0%; Pred. No. 1.3e-06;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAGTAAAG 42
    |||||||
Db 267 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAGTAAAG 226

```

## RESULT 11

```

BE818852/c
LOCUS              467 bp      mRNA      linear      EST 21-SEP-2000
DEFINITION      CM2-BN0302-100700-259-a05 BN0302 Homo sapiens cDNA, mRNA sequence.
ACCESSION      BE818852
VERSION      BE818852.1 GI:10251086
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens

```

```

REFERENCE
AUTHORS      Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V.,
O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.

```

```

TITLE      Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags

```

```

JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
PUBMED      10737800
COMMENT      Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=ct2=CM2-BN0302-100
700-259-a05&t3=2000-07-10&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 467.
Location/Qualifiers
1..467
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="BN0302"
/notes="Organ: breast normal; Vector: puc18; Site 1: SmaI;
Site 2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

```

## TITLE

```

Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags

```

```

JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
PUBMED      10737800
COMMENT      Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=ct2=CM2-BN0302-100
700-259-a05&t3=2000-07-10&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 467.
Location/Qualifiers
1..467
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="BN0302"
/notes="Organ: breast normal; Vector: puc18; Site 1: SmaI;
Site 2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

```

## JOURNAL

## PUBMED

## COMMENT

```

Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=ct2=CM2-BN0302-100
700-259-a05&t3=2000-07-10&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 467.
Location/Qualifiers
1..467
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="BN0302"
/notes="Organ: breast normal; Vector: puc18; Site 1: SmaI;
Site 2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

```

## FEATURES

## source

```

Query Match      100.0%; Score 42; DB 7; Length 467;
Best Local Similarity 100.0%; Pred. No. 1.3e-06;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

## ORIGIN

```

Query Match      100.0%; Score 42; DB 7; Length 467;
Best Local Similarity 100.0%; Pred. No. 1.3e-06;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAGTAAAG 42
    |||||||
Db 246 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAGTAAAG 205

```



RESULT 12  
 BP2533378  
 LOCUS  
 DEFINITION BP2533378 Sugano cDNA library, kidney epithelial cell Homo sapiens  
 EST. BP2533378.1 GI:52135659  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 571)  
 AUTHORS Suzuki, Y., Yamashita, R., Shirota, M., Sakakibara, Y., Chiba, J.,  
 Mizushima-Sugano, J., Nakai, K. and Sugano, S.  
 TITLE Sequence comparison of human and mouse genes reveals a homologous  
 block structure in the promoter regions  
 JOURNAL Genome Res. 14 (9), 1711-1718 (2004)  
 PUBMED 15342556  
 COMMENT Contact: Yutaka Suzuki  
 Department of Virology  
 Institute of Medical Science, University of Tokyo  
 4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
 Tel: 81-3-5449-5343  
 Fax: 81-3-5449-5416  
 Email: ysuzuki@igc.jp.  
 Location/Qualifiers  
 1..571  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="HRC03720"  
 /tissue\_type="kidney"  
 /cell\_type="epithelial cell"  
 /clone\_lib="Sugano cDNA library, kidney epithelial cell"

Query Match 100.0%; Score 42; DB 3; Length 571;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-06;  
 Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
 ||||||||||||||||||||||||||||||||||||||||||||  
 Db 522 GCCCAGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 563

RESULT 13  
 CB152636  
 LOCUS  
 DEFINITION CB152636 577 bp mRNA linear EST 29-JAN-2003  
 5', mRNA sequence.  
 ACCESSION CB152636  
 VERSION CB152636.1 GI:28137590  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 577)  
 AUTHORS Kim, N.S., Hahn, Y., Oh, J.H., Lee, J.Y., Ahn, H.Y., Chu, M.Y., Kim, M.R.,  
 Oh, K.J., Cheong, J.E., Sohn, H.Y., Kim, J.M., Park, H.S., Kim, S. and  
 Kim, Y.S.  
 TITLE 21C Frontier Korean EST Project 2001  
 JOURNAL Unpublished (2002)  
 COMMENT Contact: Kim YS  
 Genome Research Center  
 Korea Research Institute of Bioscience & Biotechnology  
 52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea  
 Tel: +82-42-860-4470  
 Fax: +82-42-860-4409

Email: yongsung@mail.kribb.re.kr  
 Plate: 5 row: G column: 07  
 High quality sequence stop: 577.  
 Location/Qualifiers  
 1..577  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="L12JSHC0s1-5-G07"  
 /sex="M"  
 /cell\_line="J-SHC"  
 /lab\_host="Top10F"  
 /clone\_lib="L12JSHC0s1"  
 /note="Organ: Liver; Vector: pCNS-D2; Site 1: EcoRI;  
 Site 2: NotI; The poly (A)+ RNA was dephosphorylated with  
 bacterial alkaline phosphatase (BAP) and then decapped  
 with tabacco acid pyrophosphatase (TAP). The decapped  
 intact mRNA was ligated with DNA-RNA linker including  
 EcoRI site by treatment of T4 RNA ligase and the first  
 strand cDNA was synthesized from oligo dT-selected mRNA by  
 priming with dT-tailed vector. The dT-tailed vector was  
 adjusted to have about 60nt. The cDNA vector was  
 circularized with E. coli DNA ligase after digestion of  
 EcoRI which site is also included in vector. An RNA strand  
 converted to a DNA strand by Okayama-Berg method. The  
 obtained cDNA vectors were used for transformation of  
 competent cells E. coli Top10F by electroporation method.  
 The cDNA libraries constructed by this method are  
 full-length enriched cDNA library. After analyzing and  
 sequencing about 2,000 - 3,000 colonies in original cDNA  
 library, the abundant cDNAs were selected and amplified by  
 PCR reaction using vector region primer including T7  
 promoter as 5' primer and N(dT)14 as 3' primer. The PCR  
 products were used as template for synthesis of  
 biotinylated single stranded RNA by in vitro transcription  
 reaction. The synthesized cDNA probes were hybridized with  
 antisense single stranded cDNAs prepared from original  
 library and incubated with avidin-gel. After removing  
 DNA-RNA hybrids by centrifuge, the subtraced cDNA  
 libraries were constructed by transformation of the  
 remaining DNA into competent cells E. coli Top10F with  
 electroporation method."

Query Match 100.0%; Score 42; DB 4; Length 577;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-06;  
 Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
 ||||||||||||||||||||||||||||||||||||||||||||  
 Db 197 GCCCAGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 238

RESULT 14  
 BP265234  
 LOCUS  
 DEFINITION BP265234 Sugano cDNA library, thyroid JTH Homo sapiens cDNA clone  
 JTH00322, mRNA sequence.  
 ACCESSION BP265234  
 VERSION BP265234.1 GI:52180465  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 580)  
 AUTHORS Suzuki, Y., Yamashita, R., Shirota, M., Sakakibara, Y., Chiba, J.,  
 Mizushima-Sugano, J., Nakai, K. and Sugano, S.  
 TITLE Sequence comparison of human and mouse genes reveals a homologous  
 block structure in the promoter regions  
 JOURNAL Genome Res. 14 (9), 1711-1718 (2004)  
 PUBMED 15342556

```

COMMENT      Contact: Yutaka Suzuki
              Department of Virology
              Institute of Medical Science, University of Tokyo
              4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
              Tel: 81-3-5449-5343
              Fax: 81-3-5449-5416
              Email: yusuzuki@hgc.jp.

FEATURES     source
              1. .580
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="JTH00322"
                /tissue_type="thyroid"
                /cell_line="JTH"
                /clone_lib="Sugano cDNA library, thyroid JTH"
                /note="thyroid tumor"

ORIGIN
Query Match      100.0%; Score 42; DB 3; Length 580;
Best Local Similarity 100.0%; Pred. No. 1.3e-06;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 42
    |||||||
Db 327 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 368

RESULT 15
LOCUS      BP268151          580 bp      mRNA      linear      EST 16-SEP-2004
DEFINITION BP268151 Sugano cDNA library, thyroid JTH Homo sapiens cDNA clone
            JTH09144, mRNA sequence.
ACCESSION  BP268151
VERSION     BP268151.1 GI:52183383
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE  1 (bases 1 to 580)
AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
JOURNAL    Genome Res. 14 (9), 1711-1718 (2004)
PUBMED    15342556
COMMENT    Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Tel: 81-3-5449-5343
            Fax: 81-3-5449-5416
            Email: yusuzuki@hgc.jp.

FEATURES     source
              1. .580
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="JTH09144"
                /tissue_type="thyroid"
                /cell_line="JTH"
                /clone_lib="Sugano cDNA library, thyroid JTH"
                /note="thyroid tumor"

ORIGIN
Query Match      100.0%; Score 42; DB 3; Length 580;
Best Local Similarity 100.0%; Pred. No. 1.3e-06;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 42
    |||||||
Db 327 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 368

RESULT 16
LOCUS      BP256144          582 bp      mRNA      linear      EST 16-SEP-2004
DEFINITION BP256144 Sugano cDNA library, kidney epithelial cell Homo sapiens
            cDNA clone HRC10671, mRNA sequence.
ACCESSION  BP256144
VERSION     BP256144.1 GI:52171374
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE  1 (bases 1 to 582)
AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
JOURNAL    Genome Res. 14 (9), 1711-1718 (2004)
PUBMED    15342556
COMMENT    Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Tel: 81-3-5449-5343
            Fax: 81-3-5449-5416
            Email: yusuzuki@hgc.jp.

FEATURES     source
              1. .582
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="HRC10671"
                /tissue_type="kidney"
                /cell_type="epithelial cell"
                /clone_lib="Sugano cDNA library, kidney epithelial cell"

ORIGIN
Query Match      100.0%; Score 42; DB 3; Length 582;
Best Local Similarity 100.0%; Pred. No. 1.3e-06;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 42
    |||||||
Db 415 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 456

RESULT 17
LOCUS      BP363758          582 bp      mRNA      linear      EST 17-SEP-2004
DEFINITION BP363758 Sugano cDNA library, fetal lung fibroblast TIG Homo
            sapiens cDNA clone TIR02383, mRNA sequence.
ACCESSION  BP363758
VERSION     BP363758.1 GI:52293963
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE  1 (bases 1 to 582)
AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
JOURNAL    Genome Res. 14 (9), 1711-1718 (2004)
PUBMED    15342556
COMMENT    Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Tel: 81-3-5449-5343
            Fax: 81-3-5449-5416
            Email: yusuzuki@hgc.jp.

FEATURES     source
              1. .580
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="JTH09144"
                /tissue_type="thyroid"
                /cell_line="JTH"
                /clone_lib="Sugano cDNA library, thyroid JTH"
                /note="thyroid tumor"

ORIGIN
Query Match      100.0%; Score 42; DB 3; Length 580;
Best Local Similarity 100.0%; Pred. No. 1.3e-06;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 42
    |||||||
Db 327 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 368

```

4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
Tel: 81-3-5449-5343  
Fax: 81-3-5449-5416  
Email: ysuzuki@hgc.jp.

# FEATURES

source  
1. .582  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="TIR02383"  
/tissue\_type="lung"  
/cell\_type="fibroblast"  
/cell\_line="TIG"  
/dev\_stage="fetal"  
/clone\_lib="Sugano cDNA library, fetal lung fibroblast TIG"

# ORIGIN

Query Match 100.0%; Score 42; DB 3; Length 582;  
Best Local Similarity 100.0%; Pred. No. 1.3e-06;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
|||||  
Db 380 GCCCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 421

# RESULT 18

BP256147  
LOCUS  
DEFINITION BP256147 Sugano cDNA library, kidney epithelial cell Homo sapiens  
ACCESSION BP256147  
VERSION BP256147.1 GI:52171377  
KEYWORDS  
SOURCE EST.  
ORGANISM Homo sapiens (human)

REFERENCE  
AUTHORS Suzuki, Y., Yamashita, R., Shirota, M., Sakakibara, Y., Chiba, J., Mizushima-Sugano, J., Nakai, K. and Sugano, S.  
TITLE Sequence comparison of human and mouse genes reveals a homologous block structure in the promoter regions  
JOURNAL Genome Res. 14 (9), 1711-1718 (2004)  
PUBMED 15342556  
COMMENT Contact: Yutaka Suzuki  
Department of Virology  
Institute of Medical Science, University of Tokyo  
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
Tel: 81-3-5449-5343  
Fax: 81-3-5449-5416  
Email: ysuzuki@hgc.jp.

# FEATURES

source  
1. .584  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="HRC10676"  
/tissue\_type="kidney"  
/cell\_type="epithelial cell"  
/clone\_lib="Sugano cDNA library, kidney epithelial cell"

# ORIGIN

Query Match 100.0%; Score 42; DB 3; Length 584;  
Best Local Similarity 100.0%; Pred. No. 1.3e-06;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
|||||  
Db 416 GCCCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 457

# RESULT 19

CB130950  
LOCUS  
DEFINITION K-EST0180946 L12JSHC0 Homo sapiens cDNA clone L12JSHC0-5-D03 5', mRNA sequence.  
ACCESSION CB130950  
VERSION CB130950.1 GI:28095412  
KEYWORDS  
SOURCE EST.  
ORGANISM Homo sapiens (human)  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
AUTHORS Kim, N.S., Hahn, Y., Oh, J.H., Lee, J.Y., Ahn, H.Y., Chu, M.Y., Kim, M.R., Oh, K.J., Cheong, J.E., Sohn, H.Y., Kim, J.M., Park, H.S., Kim, S. and Kim, Y.S.  
21C Frontier Korean EST Project 2001  
Unpublished (2002)  
Contact: Kim YS  
Genome Research Center  
Korea Research Institute of Bioscience & Biotechnology  
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea  
Tel: +82-42-860-4470  
Fax: +82-42-860-4409  
Email: yongsung@mail.kribb.re.kr  
Plate: 5 row: D column: 03  
High quality sequence stop: 621.  
Location/Qualifiers  
1. .621  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="L12JSHC0-5-D03"  
/sex="M"  
/cell\_line="J-SHC"  
/lab\_hosts="Top10F"  
/clone\_lib="L12JSHC0"  
/note="Organ: Liver; Vector: pCNS-D2; Site 1: EcoRI; Site 2: NotI; The poly (A)+ RNA was dephosphorylated with bacterial alkaline phosphatase (BAP) and then decapped with tobacco acid pyrophosphatase (TAP). The decapped intact mRNA was ligated with DNA-RNA linker including EcoRI site by treatment of T4 RNA ligase and the first strand cDNA was synthesized from oligo dt-selected mRNA by priming with dt-tailed vector. The dt-tailed vector was adjusted to have about 60nt. The cDNA vector was circularized with E. coli DNA ligase after digestion of EcoRI which site is also included in vector. An RNA strand converted to a DNA strand by Okayama-Berg method. The obtained cDNA vectors were used for transformation of competent cells E. coli Top10F by electroporation method. The cDNA libraries constructed by this method are full-length enriched cDNA library."

# TITLE

# JOURNAL

# COMMENT

# FEATURES

# source

# ORIGIN

Query Match 100.0%; Score 42; DB 4; Length 621;  
Best Local Similarity 100.0%; Pred. No. 1.3e-06;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
|||||  
Db 314 GCCCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 355

# RESULT 20

BE747107  
LOCUS  
DEFINITION BE747107 NIH\_MGC\_9 Homo sapiens cDNA clone IMAGE:3929706 5', mRNA sequence.  
ACCESSION BE747107  
VERSION BE747107.1 GI:10161099

```

KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini,
Hominiidae; Homo.
REFERENCE 1 (bases 1 to 630)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: DCTD/DTF
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: L1CM762 row: 0 column: 19
High quality sequence stop: 628.
FEATURES
source
1..630
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="S12SNU216-77-H12"
/lab_hosts="DH10B (phage-resistant)"
/clone_lib="S12SNU216"
/notes="Organ: Ovary; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

```

ORIGIN
Query Match 100.0%; Score 42; DB 7; Length 630;
Best Local Similarity 100.0%; Pred. No. 1.4e-06;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 42
|||||
Db 66 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 107
|||||

RESULT 21
BM843051
LOCUS K-EST0120649 S12SNU216 Homo sapiens cDNA clone S12SNU216-77-H12 5',
mRNA sequence.
ACCESSION BM843051
VERSION BM843051.1 GI:19199460
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.
REFERENCE 1 (bases 1 to 637)
AUTHORS Oh, J.H., Yang, J.O., Hahn, Y., Kim, M.R., Byun, S.S., Jeon, Y.J.,
Kim, J.M., Song, K.S., Noh, S.M., Kim, S., Yoo, H.S., Kim, Y.S. and
Kim, N.S.
TITLE Transcriptome analysis of human gastric cancer
JOURNAL Mamm. Genome 16 (12), 942-954 (2005)
PUBMED 16341674
COMMENT Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
```

```

FEATURES
source
1..637
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="S12SNU216-77-H12"
/sex="p"
/tissue_type="Lymph node"
/cell_type="Epithelial"
/lab_hosts="Top10F"
/clone_lib="S12SNU216"
/notes="Organ: Stomach; Vector: pcNS; Site 1: EcoRI;
Site 2: NotI; The poly (A)+ RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then decapped
with tobacco acid pyrophosphatase (TAP). The decapped
intact mRNA was ligated with DNA-RNA linker including EcoR
I site by treatment of T4 RNA ligase and the first strand
cDNA was synthesized from oligo dT-selected mRNA by
priming with dT-tailed vector. The dT-tailed vector was
adjusted to have about 60nt. The cDNA vector was
circularized with E. coli DNA ligase after digestion of
EcoRI which site is also included in vector. An RNA strand
converted to a DNA strand by Okayama-Berg method. The
obtained cDNA vectors were used for transformation of
competent cells E. coli Top10F' by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library."
```

```

ORIGIN
Query Match 100.0%; Score 42; DB 3; Length 637;
Best Local Similarity 100.0%; Pred. No. 1.4e-06;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 42
|||||
Db 59 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 100
|||||

RESULT 22
BF568108
LOCUS 602183908F1 NIH_MGC_42 Homo sapiens cDNA clone IMAGE:4300248 5',
mRNA sequence.
ACCESSION BF568108
VERSION BF568108.1 GI:11641526
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.
REFERENCE 1 (bases 1 to 699)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM158 row: 0 column: 01
High quality sequence stop: 638.
FEATURES
source
1..699
```

/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4300248"  
/tissue\_type="epithelioid carcinoma cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_42"  
/note="Organ: pancreas; Vector: pOTB7; Site 1: XhoI;  
Site 2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCACGAG(G). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by Ling  
Hong in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library. "

## ORIGIN

Query Match 100.0%; Score 42; DB 7; Length 699;  
Best Local Similarity 100.0%; Pred. No. 1.4e-06;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
|||||  
Db 321 GCCCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 362  
|||||

## RESULT 23

BG251131 702 bp mRNA linear EST 13-FEB-2001  
LOCUS 602364932F1 NIH\_MGC\_90 Homo sapiens cDNA clone IMAGE:4473439 5',  
DEFINITION mRNA sequence.  
ACCESSION BG251131  
VERSION BG251131.1 GI:12760947  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
1 (bases 1 to 702)  
NIH-MGC http://mgc.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LHAM10295 row: 0 column: 08  
High quality sequence stop: 688.  
Location/Qualifiers

## FEATURES

source 1..702  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4473439"  
/tissue\_type="adenocarcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_90"  
/note="Organ: liver; Vector: pCMV-SPORT6; Site 1: NotI;  
Site 2: SalI; Cloned unidirectionally; oligo-dT primed.  
Average insert size 1.7 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 100.0%; Score 42; DB 2; Length 702;  
Best Local Similarity 100.0%; Pred. No. 1.4e-06;

Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
|||||  
Db 464 GCCCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 505  
|||||

## RESULT 24

BQ887352 851 bp mRNA linear EST 16-AUG-2002  
LOCUS AGENCOURT 8678064 NIH\_MGC\_40 Homo sapiens cDNA clone IMAGE:6381080  
DEFINITION 5', mRNA sequence.  
ACCESSION BQ887352  
VERSION BQ887352.1 GI:22279366  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
1 (bases 1 to 851)  
NIH-MGC http://mgc.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: DCTD/DTF  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LUCM2570 row: 1 column: 09  
High quality sequence stop: 579.  
Location/Qualifiers

## REFERENCE

AUTHORS NIH-MGC  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: DCTD/DTF  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LUCM2570 row: 1 column: 09  
High quality sequence stop: 579.  
Location/Qualifiers

## FEATURES

source 1..851  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6381080"  
/tissue\_type="carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_40"  
/note="Organ: prostate; Vector: pOTB7; Site 1: XhoI;  
Site 2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCACGAG(G). Library constructed by  
Ling Hong in the laboratory of Gerald M. Rubin (University  
of California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 100.0%; Score 42; DB 3; Length 851;  
Best Local Similarity 100.0%; Pred. No. 1.4e-06;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## Qy 1 GCCCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42

|||||  
Db 182 GCCCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 223  
|||||

## RESULT 25

CA454746 865 bp mRNA linear EST 12-NOV-2002  
LOCUS AGENCOURT 10763191 MAPCL Homo sapiens cDNA clone IMAGE:6721245 5',  
DEFINITION mRNA sequence.  
ACCESSION CA454746  
VERSION CA454746.1 GI:24904781  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.

#### REFERENCE AUTHORS TITLE JOURNAL COMMENT

1 (bases 1 to 865)  
NIH-MGC <http://mgs.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: [cgaps-remail.nih.gov](mailto:cgaps-remail.nih.gov)

Tissue Procurement: Kristi A. Eglund, Ira Pastan

cDNA Library Preparation: Invitrogen Corp

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: L14M14282 row: a column: 21

High quality sequence stop: 886.

#### FEATURES source

1..865  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6721245"  
/cell\_line="ZR-75-1, MCF7, SK-BR-3, MDA-MB-231,  
HRT-HME1, LNCaP"  
/lab\_host="EMDH10B"  
/clone\_lib="NAPCL"  
/note="Vector: pCMV-SPORT6; Site 1: EcoRV; Site 2: Not I;  
Subtracted with brain, liver, lung, kidney and muscle.  
Directionally cloned. Priming method: oligo-dT. Average  
insert size: 1800 bp. Library amplification: 26,000 fold.  
Kristi A. Eglund, James J. Vincent, Robert Strausberg,  
Bungkook Lee & Ira Pastan. Discovery of new breast  
cancer genes encoding membrane and secreted proteins.  
Manuscript submitted."

#### ORIGIN

Query Match 100.0%; Score 42; DB 4; Length 865;  
Best Local Similarity 100.0%; Pred. No. 1.4e-06;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 42

Db 611 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 652

#### RESULT 26 BM011054 LOCUS

DEFINITION 603634744F1 NIH\_MGC\_47 Homo sapiens cDNA clone IMAGE:5421755 5',  
mRNA sequence.

ACCESSION BM011054

VERSION BM011054.1 GI:16525408

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.

1 (bases 1 to 877)

NIH-MGC <http://mgs.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: [cgaps-remail.nih.gov](mailto:cgaps-remail.nih.gov)

Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: L14M1878 row: h column: 12

High quality sequence stop: 746.

#### FEATURES source

Location/Qualifiers  
1..877  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5421755"  
/tissue\_type="neuroblastoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_47"  
/note="Organ: brain; Vector: pOTB7; Site 1: XhoI; Site 2:  
EcoRI; cDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'  
adaptor: GGCACGAG(G). Size-selected >500bp for average  
insert size 1.8kb. Library constructed by Ling Hong in  
the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library."

#### ORIGIN

Query Match 100.0%; Score 42; DB 2; Length 877;

Best Local Similarity 100.0%; Pred. No. 1.4e-06;

Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 42

Db 184 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 225

#### RESULT 27

BQ958341

LOCUS

DEFINITION AGENCOURT 10037048 NIH\_MGC\_40 Homo sapiens cDNA clone IMAGE:6480481

5' RNA sequence.

ACCESSION BQ958341

VERSION BQ958341.1 GI:22373819

KEYWORDS EST.

SOURCE

ORGANISM

Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.

1 (bases 1 to 899)

NIH-MGC <http://mgs.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: [cgaps-remail.nih.gov](mailto:cgaps-remail.nih.gov)

Tissue Procurement: DCTD/DTF

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: L14M2660 row: f column: 02

High quality sequence stop: 636.

#### FEATURES source

Location/Qualifiers  
1..899  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6480481"  
/tissue\_type="carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_40"  
/note="Organ: prostate; Vector: pOTB7; Site 1: XhoI;  
Site 2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCACGAG(G). Library constructed by  
Ling Hong in the laboratory of Gerald M. Rubin (University

of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 100.0%; Score 42; DB 3; Length 899;  
Best Local Similarity 100.0%; Pred. No. 1.4e-06;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
|||||  
Db 546 GCCCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 587  
|||||

## RESULT 28

BG831564  
LOCUS 602765819f1 NIH\_MGC\_42 Homo sapiens cDNA clone IMAGE:4907885 5',  
DEFINITION mRNA sequence.

ACCESSION BG831564

VERSION BG831564.1 GI:14179151

KEYWORDS EST.

SOURCE Homo sapiens (human)

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.

1 (bases 1 to 901)

NIH-MGC http://mgs.nci.nih.gov/.

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgabs-r@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLCM1813 row: e column: 06

High quality sequence stop: 663.

## FEATURES

Location/Qualifiers

1..901

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:4907885"

/tissue type="epithelioid carcinoma cell line"

/lab host="DH10B (phage-resistant)"

/clone.lib="NIH\_MGC\_42"

/note="Organ: pancreas; Vector: pOTB7; Site\_1: XhoI;

Site\_2: EcoRI; cDNA made by oligo-dT priming.

Directionally cloned into EcoRI/XhoI sites using the

following 5' adaptor: GGCACGAG(G). Size-selected >500bp

for average insert size 1.8kb. Library constructed by Ling

Hong in the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies).

Note: this is a NIH\_MGC Library. |"

## ORIGIN

Query Match 100.0%; Score 42; DB 2; Length 901;  
Best Local Similarity 100.0%; Pred. No. 1.4e-06;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
|||||

Db 322 GCCCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 363  
|||||

## RESULT 29

BX372160/c

## LOCUS

DEFINITION BX372160 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens  
cDNA clone CS0DK011YF17 3-PRIME, mRNA sequence.

## ACCESSION

VERSION BX372160

KEYWORDS BX372160.1 GI:30452082

SOURCE EST.

ORGANISM Homo sapiens (human)

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.

1 (bases 1 to 918)

Li.W.B., Gruber.C., Jessee,J. and Polayes,D.

Full-length cDNA libraries and normalization

Unpublished (2001)

Contact: Genoscope

Genoscope - Centre National de Sequencage

2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE

Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime

end enriched, double-strand cDNA was digested with Not I and cloned

into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library

was normalized. Library was constructed by Life Technologies, a

division of Invitrogen. This sequence belongs to sequence cluster

6601.f

For more information about this cluster, see

http://www.genoscope.cns.fr/cdna?s=CS0BA1040ZB07\_CS03797\_1&c=6601.f

## FEATURES

source

Location/Qualifiers

1..918

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clones="CS0DK011YF17"

/cell\_type="HELA CELLS COT 25-NORMALIZED"

/clone.lib="HELA"

/note="1st strand cDNA was primed with a NotI-oligo(dT)

primer. Five prime end enriched, double-strand cDNA was

digested with Not I and cloned into the Not I and EcoR V

sites of the pCMVSPORT 6 vector. Library was normalized."

## ORIGIN

Query Match 100.0%; Score 42; DB 4; Length 918;

Best Local Similarity 100.0%; Pred. No. 1.5e-06;

Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
|||||

Db 335 GCCCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 294  
|||||

## RESULT 30

BUS43952

LOCUS

DEFINITION AGNCOURT 10133041 NIH\_MGC\_40 Homo sapiens cDNA clone IMAGE:6576152  
5', mRNA sequence.

ACCESSION BUS43952

VERSION BUS43952.1 GI:22854435

KEYWORDS EST.

SOURCE Homo sapiens (human)

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.

1 (bases 1 to 923)

NIH-MGC http://mgs.nci.nih.gov/.

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgabs-r@mail.nih.gov

Tissue Procurement: DCTD/DTF

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLCM2774 row: d column: 08  
 High quality sequence stop: 641.  
 Location/Qualifiers  
 1. 923  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6576152"  
 /tissue\_type="carcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC 40"  
 /note="Organ: prostate; Vector: pOTB7; Site 1: XhoI;  
 Site 2: EcoRI; cDNA made by oligo-dT priming.  
 Directionally cloned into EcoRI/XhoI sites using the  
 following 5' adaptor: GCACGAG(G). Library constructed by  
 Ling Hong in the laboratory of Gerald M. Rubin (University  
 of California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies).  
 Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 100.0%; Score 42; DB 3; Length 923;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-06;  
 Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCGAGCCCTTACAAAGACTATGAGTAAAG 42  
 |||||  
 Db 546 GCCCAGCGTTGGCGAGCCCTTACAAAGACTATGAGTAAAG 587

## RESULT 31

BX370558  
 LOCUS BX370558 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens  
 DEFINITION cDNA clone CS0DK011YF17 5-PRIME, mRNA sequence.  
 ACCESSION BX370558  
 VERSION BX370558.1 GI:30453927  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)

## ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homnidae; Homo.

## REFERENCE

1 (bases 1 to 962)  
 Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
 Full-length cDNA libraries and normalization  
 Unpublished (2001)

## AUTHORS

Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
 Full-length cDNA libraries and normalization  
 Unpublished (2001)  
 Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE  
 Email: [segre@genoscope.cns.fr](mailto:segre@genoscope.cns.fr), Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
 end enriched, double-strand cDNA was digested with Not I and cloned  
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
 was normalized. Library was constructed by Life Technologies, a  
 division of Invitrogen. This sequence belongs to sequence cluster  
 6601.f

For more information about this cluster, see  
[http://www.genoscope.cns.fr/cdna?s=CS0BAG0492C03\\_CS04627\\_lkc=6601.f](http://www.genoscope.cns.fr/cdna?s=CS0BAG0492C03_CS04627_lkc=6601.f)

## FEATURES

source  
 Location/Qualifiers  
 1. 962  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0DK011YF17"  
 /cell\_type="HELA CELLS COT 25-NORMALIZED"  
 /cell\_line="HELA"

/clone\_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-oligo(dT)  
 primer. Five prime end enriched, double-strand cDNA was  
 digested with Not I and cloned into the Not I and EcoR V  
 sites of the pCMVSPORT 6 vector. Library was normalized."

## ORIGIN

Query Match 100.0%; Score 42; DB 4; Length 962;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-06;  
 Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCGAGCCCTTACAAAGACTATGAGTAAAG 42  
 |||||  
 Db 248 GCCCAGCGTTGGCGAGCCCTTACAAAGACTATGAGTAAAG 289

## RESULT 32

BQ067433  
 LOCUS BQ067433 1023 bp mRNA linear EST 02-APR-2002  
 DEFINITION AGENCOURT\_6758944 NIH\_MGC\_115 Homo sapiens cDNA clone IMAGE:5754944  
 5', mRNA sequence.  
 ACCESSION BQ067433  
 VERSION BQ067433.1 GI:19896479  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)

## ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homnidae; Homo.

## REFERENCE

1 (bases 1 to 1023)  
 NIH-MGC <http://mgs.nci.nih.gov/>.

## AUTHORS

National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-x@mail.nih.gov](mailto:cgapbs-x@mail.nih.gov)

## TITLE

Tissue Procurement: Life Technologies, Inc.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: L1AM12793 row: c column: 09

High quality sequence stop: 649.

## FEATURES

source  
 Location/Qualifiers  
 1. 1023  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5754944"  
 /lab\_host="DH10B"  
 /clone\_lib="NIH\_MGC\_115"  
 /note="Organ: pooled brain, lung, testis; Vector:  
 pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA  
 source anonymous pool of 6 male brains, age range 23-27; 1  
 male lung, age 27; and 1 male testis, age 69. Library is  
 oligo-dT primed and directionally cloned (EcoRV site is  
 destroyed upon cloning). Average insert size 1.8 kb,  
 insert size range 1-3 kb. Library is normalized and  
 enriched for full-length clones and was constructed by C.  
 Gruber (Invitrogen). Research Genetics tracking code  
 021. Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 100.0%; Score 42; DB 3; Length 1023;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-06;  
 Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCGAGCCCTTACAAAGACTATGAGTAAAG 42  
 |||||  
 Db 403 GCCCAGCGTTGGCGAGCCCTTACAAAGACTATGAGTAAAG 444



```
RESULT 33
CR597125      1830 bp      mRNA      linear      HTC 21-JUL-2004
LOCUS
DEFINITION   full-length cDNA clone CS0DK011YF17 of HeLa cells Cot 25-normalized
              of Homo sapiens (human).
ACCESSION   CR597125
VERSION     CR597125.1 GI:50477932
KEYWORDS    HTC; CNSLT cDNA.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homnidae; Homo.
REFERENCE   1 (bases 1 to 1830)
AUTHORS    Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE      Full-length cDNA libraries and normalization
JOURNAL    Unpublished
REMARK     Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InvitroGen Corporation 1600
            Faraday Avenue
            Genoscope.
REFERENCE   2 (bases 1 to 1830)
AUTHORS
TITLE      Direct Submission
JOURNAL    Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
FEATURES
source
1..1830
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DK011YF17"
/tissue type="HeLa cells Cot 25-normalized"
/plasmid="pCMVSPORT_6"

Query Match      100.0%; Score 42; DB 6; Length 1830;
Best Local Similarity 100.0%; Pred. No. 1.6e-06;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42
|||||
Db 1118 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1159

RESULT 34
CF780868      916 bp      mRNA      linear      EST 20-OCT-2003
LOCUS
DEFINITION   AGNCOURT_15939245 NIH_MGC_219 Homo sapiens cDNA clone
              IMAGE:30523569 5', mRNA sequence.
ACCESSION   CF780868
VERSION     CF780868.1 GI:37740645
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homnidae; Homo.
REFERENCE   1 (bases 1 to 916)
AUTHORS    NIH-MGC http://mgs.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL    Unpublished (1999)
COMMENT     Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgabs-r@mail.nih.gov
            Tissue Procurement: James Martin, University of Iowa
```

```
CDNA Library Preparation: M. Bento Soares, University of Iowa
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: NDAM601 row: 1 column: 10
High quality sequence stop: 651.

FEATURES
source
1..916
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30523569"
/tissue type="Pooled Chondrosarcoma Tumor cells"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC_219"
/note="Vector: pYX-Asc; Site1: EcoRI; Site2: NotI;
Library is oligo-dT primed and directionally cloned
Denatured RNA was size fractionated on a 1% agarose gel.
First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated with
EcoR I adaptor, digested with Not I and then cloned
directionally into pYX-Asc vector. Average insert size
0.5-1Kb. Adaptors 5' (AATTGCGCAGG)3' and 5'd
(CCTCGGCCG)3'. 3' linker sequence - GCGGCGCTGAGAGCC T18.
Sequencing primers 3' end: T3 promoter primer 5'd
(ATTAACCTCCTACTAAGG)3'. 5' End: T7 promoter primer 5'd
(TAATGACTACTATAGG)3'. Library was constructed in the
laboratory of M. Bento Soares. Average insert size 2-3kb.
Note: this is a NIH_MGC Library."

ORIGIN
Query Match      97.6%; Score 41; DB 5; Length 916;
Best Local Similarity 97.6%; Pred. No. 3.9e-06;
Matches 41; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42
|||||
Db 667 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 708

RESULT 35
BI193620      1125 bp      mRNA      linear      EST 10-JUL-2001
LOCUS
DEFINITION   602946519F1 NIH_MGC_42 Homo sapiens cDNA clone IMAGE:5089900 5',
              mRNA sequence.
ACCESSION   BI193620
VERSION     BI193620.1 GI:14648640
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homnidae; Homo.
REFERENCE   1 (bases 1 to 1125)
AUTHORS    NIH-MGC http://mgs.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL    Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgabs-r@mail.nih.gov
            Tissue Procurement: ATCC
            CDNA Library Preparation: Ling Hong/Rubin Laboratory
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1851 row: e column: 05
            High quality sequence stop: 529.

FEATURES
source
1..1125
```

/organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5089900"  
 /tissue\_type="epithelioid carcinoma cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_42"  
 /note="Organ: pancreas; Vector: pOTB7; Site 1: XhoI;  
 Site 2: EcoRI; cDNA made by oligo-dT priming.  
 Directionally cloned into EcoRI/XhoI sites using the  
 following 5' adaptor: GGCACGAG(G). Size-selected >500bp  
 for average insert size 1.8kb. Library constructed by Ling  
 Hong in the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies).  
 Note: this is a NIH\_MGC Library. |"

## ORIGIN

Query Match 97.6%; Score 41; DB 2; Length 1125;  
 Best Local Similarity 100.0%; Pred. No. 4e-06;  
 Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGCGCGAGCCCTTACAAAGACTATGAAGTAAA 41  
 |||||  
 Db 294 GCCACGCTTGCGCGAGCCCTTACAAAGACTATGAAGTAAA 334

RESULT 36  
BE818831/c

LOCUS BE818831 365 bp mRNA linear EST 21-SEP-2000  
 DEFINITION CM2-BN0302-050700-256-al2 BN0302 Homo sapiens cDNA, mRNA sequence.

ACCESSION BE818831  
 VERSION BE818831.1 GI:10251065  
 KEYWORDS EST.

SOURCE Homo sapiens (human)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Hominiidae; Homo.

REFERENCE 1 (bases 1 to 365)

AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,  
 Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V.,  
 O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
 Simpson, A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

PUBMED 10737800

COMMENT Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome  
 Project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-CM2-BN0302-050  
 700-256-al2&t3=2000-07-05&t4=1)

Seq primer: puc 18 forward

High quality sequence start: 14

High quality sequence stop: 365.

## FEATURES

source

Location/Qualifiers  
 1..365

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/dev\_stage="Adult"

/clone\_lib="BN0302"

/note="Organ: breast\_normal; Vector: puc18; Site 1: SmaI;

Site 2: SmaI; A mini-library was made by cloning products  
 derived from ORESTES PCR (U.S. Letters Patent application  
 No. 196,716 - Ludwig Institute for Cancer Research)  
 profiles into the pUC 18 vector. Reverse transcription of  
 tissue mRNA and cDNA amplification were performed under  
 low stringency conditions."

## ORIGIN

Query Match 96.2%; Score 40.4; DB 7; Length 365;  
 Best Local Similarity 97.6%; Pred. No. 5.9e-06;  
 Matches 41; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGCGCGAGCCCTTACAAAGACTATGAAGTAAA 42  
 |||||  
 Db 267 GCCACGCTTGCGCGAGCCCTTACAAAGACTATGAAGTAAA 226

## RESULT 37

BE818807/c

LOCUS BE818807 449 bp mRNA linear EST 21-SEP-2000  
 DEFINITION CM2-BN0302-040700-253-d02 BN0302 Homo sapiens cDNA, mRNA sequence.

ACCESSION BE818807

VERSION BE818807.1 GI:10251041

KEYWORDS EST.

SOURCE Homo sapiens (human)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Hominiidae; Homo.

REFERENCE 1 (bases 1 to 449)

AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,  
 Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V.,  
 O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
 Simpson, A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

PUBMED 10737800

COMMENT Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome  
 Project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-CM2-BN0302-040  
 700-253-d02&t3=2000-07-04&t4=1)

Seq primer: puc 18 forward

High quality sequence start: 24

High quality sequence stop: 306.

Location/Qualifiers

## FEATURES

source

1..449

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/dev\_stage="Adult"

/clone\_lib="BN0302"

/note="Organ: breast\_normal; Vector: puc18; Site 1: SmaI;  
 Site 2: SmaI; A mini-library was made by cloning products  
 derived from ORESTES PCR (U.S. Letters Patent application  
 No. 196,716 - Ludwig Institute for Cancer Research)  
 profiles into the pUC 18 vector. Reverse transcription of  
 tissue mRNA and cDNA amplification were performed under  
 low stringency conditions."

## ORIGIN

Query Match 96.2%; Score 40.4; DB 7; Length 449;  
 Best Local Similarity 97.6%; Pred. No. 6.1e-06;

```
Matches 41; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 GCCCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42
|||||
Db 240 GCCCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 199
|||||

RESULT 38
BP267627 580 bp mRNA linear EST 16-SEP-2004
LOCUS BP267627 Sugano cDNA library, thyroid JTH Homo sapiens cDNA clone
DEFINITION JTH07584, mRNA sequence.
ACCESSION BP267627
VERSION BP267627.1 GI:52182859
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 580)
AUTHORS Suzuki, Y., Yamashita, R., Shirota, M., Sakakibara, Y., Chiba, J.,
Mizushima-Sugano, J., Nakai, K. and Sugano, S.
TITLE Sequence comparison of human and mouse genes reveals a homologous
block structure in the promoter regions
JOURNAL Genome Res. 14 (9), 1711-1718 (2004)
PUBMED 15342556
COMMENT Contact: Yutaka Suzuki
Department of Virology
Institute of Medical Science, University of Tokyo
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
Tel: 81-3-5449-5343
Fax: 81-3-5449-5416
Email: ysuzuki@hgc.jp.
Location/Qualifiers
1 .580
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="JTH07584"
/tissue_type="thyroid"
/cell_line="JTH"
/clone_lib="Sugano cDNA library, thyroid JTH"
/note="thyroid tumor"

FEATURES
source
1 .580
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="JTH07584"
/tissue_type="thyroid"
/cell_line="JTH"
/clone_lib="Sugano cDNA library, thyroid JTH"
/note="thyroid tumor"

ORIGIN
Query Match 96.2%; Score 40.4; DB 3; Length 580;
Best Local Similarity 97.6%; Pred. No. 6.4e-06;
Matches 41; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 GCCCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42
|||||
Db 539 GCCCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 580
|||||

RESULT 39
BP685729 940 bp mRNA linear EST 15-JUL-2002
LOCUS BP685729 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6251057
DEFINITION AGNCOURT_8344399 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6251057
5', mRNA sequence.
ACCESSION BQ685729
VERSION BQ685729.1 GI:21811045
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 940)
AUTHORS NIH-MGC Http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
```

```
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM2395 row: b column: 18
High quality sequence stop: 538.
Location/Qualifiers
1 .940
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6251057"
/tissue_type="ductal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_110"
/note="Organ: pancreas; Vector: pOTB7; Site 1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dT priming
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GCCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

ORIGIN
Query Match 96.2%; Score 40.4; DB 3; Length 940;
Best Local Similarity 97.6%; Pred. No. 7e-06;
Matches 41; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 GCCCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42
|||||
Db 687 GCCCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 728
|||||

RESULT 40
AUI40676 680 bp mRNA linear EST 05-AUG-2002
LOCUS AUI40676 PLACE4 Homo sapiens cDNA clone PLACE4000034 5', mRNA
DEFINITION sequence.
ACCESSION AUI40676
VERSION AUI40676.1 GI:11002197
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 680)
AUTHORS Kimura, K., Wakamatsu, A., Suzuki, Y., Ota, T., Nishikawa, T.,
Yamashita, R., Yamamoto, J., Sekine, M., Tsuritani, K., Wakaguri, H.,
Ishii, S., Sugiyama, T., Saito, K., Isono, Y., Irie, R., Kushida, N.,
Yoneyama, T., Otsuka, R., Kanda, K., Yokoi, T., Kondo, H., Wagatsuma, M.,
Murakawa, K., Iehida, S., Ishibashi, T., Takahashi-Fujii, A.,
Tanase, T., Nagai, K., Kikuchi, H., Nakai, K., Isogai, T. and Sugano, S.
TITLE Diversification of Transcriptional Modulation: Large-scale
Identification and Characterization of Putative Alternative
Promoters of Human Genes
JOURNAL Genome Res. 16 (1), 55-65 (2006)
PUBMED 16344560
COMMENT Contact: Takao Isogai
FLJ Project (HRI Team)
Helix Research Institute
2-6-7 Kazusa-Kamatari, Kisarazu, Chiba, 292-0818, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: flj-cdna@nifty.com
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
```

```
Helix Research Institute.
Location/Qualifiers
1. .680
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="PLACE4000034"
/tissue_type="placenta"
/clone_lib="PLACE4"
/note="Vector: pME18SFL3"

FEATURES
source
Query Match 80.0%; Score 33.6; DB 1; Length 680;
Best Local Similarity 87.8%; Pred. No. 0.0051;
Matches 36; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAACTAAA 41
Db 633 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAAAGTAA 673

ORIGIN
Query Match 80.0%; Score 33.6; DB 1; Length 680;
Best Local Similarity 87.8%; Pred. No. 0.0051;
Matches 36; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

RESULT 41
LOCUS AAL12374 367 bp mRNA linear EST 01-AUG-1997
DEFINITION z68e10.11 Stratagene HeLa cell s3 937216 Homo sapiens cDNA clone
ACCESSION AAL12374
VERSION AAL12374.1 GI:1664784
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 367)
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
Chissoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, N.,
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,
Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,
Trevas, E., Underwood, K., Wohldmann, P., Waterston, R., Wilson, R.
and Marra, M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
8889549
Contact: wilson@wustl.edu
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1363 Std Error: 0.00
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 329.
Location/Qualifiers
1. .367
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:4593181"
/db_xref="taxon:9606"
/clone="IMAGE:563370"
/sex="female"
/dev_stage="HeLa S3 cell line"
/lab_host="SOJR (kanamycin resistant)"
/clone_lib="Stragene HeLa cell s3 937216"
/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
XhoI; Cloned unidirectionally. Primer: Oligo dt. HeLa S3
epithelioid carcinoma cells grown to semi-confluency
without induction. Average insert size: 1.5 kb; Uni-ZAP XR
Vector -5' adaptor sequence: 5' GAATTCGGCAG 3' -3'
adaptor sequence: 5' CTCGAGTTTTTTTTTTTTT 3'"

FEATURES
source
Query Match 72.4%; Score 30.4; DB 7; Length 583;
Best Local Similarity 96.9%; Pred. No. 0.11;
Matches 31; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAAAGTAA 32
Db 552 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAAAGTAA 583

ORIGIN
Query Match 72.4%; Score 30.4; DB 7; Length 583;
Best Local Similarity 96.9%; Pred. No. 0.11;
Matches 31; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAAAGTAA 42
Db 116 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAAAGTAA 158

ORIGIN
Query Match 73.8%; Score 31; DB 1; Length 367;
Best Local Similarity 97.7%; Pred. No. 0.058;
Matches 42; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAAAGTAA 42
Db 116 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAAAGTAA 158

ORIGIN
Query Match 73.8%; Score 31; DB 1; Length 367;
Best Local Similarity 97.7%; Pred. No. 0.058;
Matches 42; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

RESULT 42
LOCUS BE006062 583 bp mRNA linear EST 05-JUN-2000
DEFINITION RC2-BN0123-170300-011-c07 BN0123 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE006062
VERSION BE006062.1 GI:8266295
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 583)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
Brunstein, A., deOliveira, P.S., Bucher, P., Jorgensen, C.V.,
O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
10737800
Contact: simpson@ludwig.org.br
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2=RC2-BN0123-170
300-011-c07&t3=2000-03-17&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 583.
Location/Qualifiers
1. .583
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="BN0123"
/note="Organ: breast normal; Vector: puc18; Site 1: SmaI;
Site 2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
```

```

RESULT 43
DN875318
LOCUS      DN875318          606 bp      mRNA      linear      EST 21-APR-2005
DEFINITION cdna clone n3d38908.Y1 Dog eye cornea. Unnormalized (nad) Canis familiaris
ACCESSION  DN875318
VERSION     DN875318.1  GI:62845259
KEYWORDS   EST.
SOURCE     Canis familiaris (dog)
ORGANISM   Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
            Canis.
REFERENCE  1  (bases 1 to 606)
AUTHORS   Wistow,G.
TITLE     NEIBank analysis of Dog cornea
JOURNAL   Unpublished (2005)
COMMENT   Contact: Wistow G
            Section on Molecular Structure and Function
            National Eye Institute
            6/331, NIH, Bethesda, MD 20892-2740, USA
            Tel: 301 402 3452
            Fax: 301 496 0078
            Email: graeme@helix.nih.gov
            Plate: 38 row: g column: 08
            Seq primer: Universal M13 Reverse.
            Location/Qualifiers
FEATURES   source
            1..606
                /organism="Canis familiaris"
                /mol_type="mRNA"
                /strain="Beagle"
                /db_xref="taxon:9615"
                /clone="nad38908"
                /tissue_type="Cornea"
                /dev_stage="Adult"
                /lab_host="EMDH10B"
                /clone_lib="Dog eye cornea. Unnormalized (nad)"
                /note="Organ: Eye; Vector: pCMVSPORT6; RNA was extracted
                from dog cornea tissue. A directionally cloned cDNA
                library in the pCMVSPORT6 vector (Invitrogen) was
                constructed at Bioserve Biotechnology (Laurel MD)
                essentially following the protocols of the SuperScript
                Plasmid System, full details of which are contained in the
                manufacturer's instruction manual
                (http://www.lifetech.com/). First strand synthesis was
                carried out using a Not I primer-adaptor
                [5'-pGACTAGTTCTAGATCGGCGGCCGCTT15-3']. cDNA was
                cloned in Not I/Sal I sites. EST analysis was performed at
                the NIH Intramural Sequencing Center (NISC). Analyzed data
                available through http://neibank.nei.nih.gov."
ORIGIN
Query Match      69.5%; Score 29.2; DB 9; Length 606;
Best Local Similarity 81.0%; Pred. No. 0.37;
Matches 34; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy  1  GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   286 GCCCAGACTCGGACTAGCCCTTACAAATGACTATGAAGGAGG 327

RESULT 44
BE548141
LOCUS      BE548141          912 bp      mRNA      linear      EST 09-AUG-2000
DEFINITION 601072396F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:3458567 5',
            mRNA sequence.
ACCESSION  BE548141
VERSION     BE548141.1  GI:9776786
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE  1  (bases 1 to 912)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: Incyte Genomics, Inc.
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM8449 row: p column: 24
            High quality sequence stop: 651.
            Location/Qualifiers
FEATURES   source
            1..912
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:3458567"
                /tissue_type="cervical carcinoma cell line"
                /lab_host="DH10B"
                /clone_lib="NIH_MGC_12"
                /note="Organ: Cervix; Vector: pCMV-SPORT6; Site 1: NotI;
                Site 2: SalI; Cloned unidirectionally. Primer: Oligo dr.
                Average insert size 1.4 kb. Library prepared by Life
                Technologies."
ORIGIN
Query Match      67.6%; Score 28.4; DB 7; Length 912;
Best Local Similarity 82.1%; Pred. No. 0.88;
Matches 32; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy  4  CACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   620 CCCACGCGGCGAGCCCTTACAAAGACTATGAAGTCAAG 658

RESULT 45
AW390096
LOCUS      AW390096          528 bp      mRNA      linear      EST 04-FEB-2000
DEFINITION CM3-ST0179-051099-019-f07 ST0179 Homo sapiens cDNA, mRNA sequence.
ACCESSION  AW390096
VERSION     AW390096.1  GI:6894755
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE  1  (bases 1 to 528)
AUTHORS   HCGP http://www.ludwig.org.br/ORESTES.
TITLE     The FAPESP/LICR Human Cancer Genome Project
JOURNAL   Unpublished (1999)
COMMENT   Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM3&t2=CM3-ST0179-
            051099-019-f07&t3=1999-10-05&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 18
            High quality sequence stop: 528.
            Location/Qualifiers
FEATURES

```

```

REFERENCE
AUTHORS   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
TITLE     1  (bases 1 to 912)
JOURNAL   NIH-MGC http://mgc.nci.nih.gov/.
COMMENT   National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: Incyte Genomics, Inc.
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM8449 row: p column: 24
            High quality sequence stop: 651.
            Location/Qualifiers
FEATURES   source
            1..912
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:3458567"
                /tissue_type="cervical carcinoma cell line"
                /lab_host="DH10B"
                /clone_lib="NIH_MGC_12"
                /note="Organ: Cervix; Vector: pCMV-SPORT6; Site 1: NotI;
                Site 2: SalI; Cloned unidirectionally. Primer: Oligo dr.
                Average insert size 1.4 kb. Library prepared by Life
                Technologies."
ORIGIN
Query Match      67.6%; Score 28.4; DB 7; Length 912;
Best Local Similarity 82.1%; Pred. No. 0.88;
Matches 32; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy  4  CACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   620 CCCACGCGGCGAGCCCTTACAAAGACTATGAAGTCAAG 658

RESULT 45
AW390096
LOCUS      AW390096          528 bp      mRNA      linear      EST 04-FEB-2000
DEFINITION CM3-ST0179-051099-019-f07 ST0179 Homo sapiens cDNA, mRNA sequence.
ACCESSION  AW390096
VERSION     AW390096.1  GI:6894755
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE  1  (bases 1 to 528)
AUTHORS   HCGP http://www.ludwig.org.br/ORESTES.
TITLE     The FAPESP/LICR Human Cancer Genome Project
JOURNAL   Unpublished (1999)
COMMENT   Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM3&t2=CM3-ST0179-
            051099-019-f07&t3=1999-10-05&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 18
            High quality sequence stop: 528.
            Location/Qualifiers
FEATURES

```

```

source
1. 528
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="ST0179"
/notes="Organ: stomach; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

ORIGIN
Query Match 64.3%; Score 27; DB 7; Length 528;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGCGGAGCCCTTACAAA 27
| | | | | | | | | | | | | | | | | | | | |
Db 502 GCCACGCTTGCGGAGCCCTTACAAA 528

RESULT 46
BP267918
LOCUS
DEFINITION BP267918 Sugano cDNA library, thyroid JTH Homo sapiens cDNA clone
JTH08419, mRNA sequence.
ACCESSION BP267918
VERSION BP267918.1 GI:52183150
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1 (bases 1 to 582)
Suzuki,Y., Yamashita,R., Shiota,M., Sakakibara,Y., Chiba,J.,
Mizushima-Sugano,J., Nakai,K. and Sugano,S.
Sequence comparison of human and mouse genes reveals a homologous
block structure in the promoter regions
Genome Res. 14 (9), 1711-1718 (2004)
Contact: Yutaka Suzuki
Department of Virology
Institute of Medical Science, University of Tokyo
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
Tel: 81-3-5449-5343
Fax: 81-3-5449-5416
Email: ysuzuki@hgc.jp.
Location/Qualifiers
1. 582
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="JTH08419"
/tissue_type="thyroid"
/cell_line="JTH"
/clone_lib="Sugano cDNA library, thyroid JTH"
/notes="thyroid tumor"

```

```

ORIGIN
Query Match 64.3%; Score 27; DB 3; Length 582;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGCGGAGCCCTTACAAA 27
| | | | | | | | | | | | | | | | | | | | |
Db 556 GCCACGCTTGCGGAGCCCTTACAAA 582

RESULT 47
BP267918
LOCUS
DEFINITION BP267918 Sugano cDNA library, thyroid JTH Homo sapiens cDNA clone
JTH08419, mRNA sequence.
ACCESSION BP267918
VERSION BP267918.1 GI:52183150
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1 (bases 1 to 582)
Suzuki,Y., Yamashita,R., Shiota,M., Sakakibara,Y., Chiba,J.,
Mizushima-Sugano,J., Nakai,K. and Sugano,S.
Sequence comparison of human and mouse genes reveals a homologous
block structure in the promoter regions
Genome Res. 14 (9), 1711-1718 (2004)
Contact: Yutaka Suzuki
Department of Virology
Institute of Medical Science, University of Tokyo
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
Tel: 81-3-5449-5343
Fax: 81-3-5449-5416
Email: ysuzuki@hgc.jp.
Location/Qualifiers
1. 582
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="JTH08419"
/tissue_type="thyroid"
/cell_line="JTH"
/clone_lib="Sugano cDNA library, thyroid JTH"
/notes="thyroid tumor"

```

```

ORIGIN
Query Match 64.3%; Score 27; DB 3; Length 582;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGCGGAGCCCTTACAAA 27
| | | | | | | | | | | | | | | | | | | | |
Db 556 GCCACGCTTGCGGAGCCCTTACAAA 582

RESULT 47
BP267918
LOCUS
DEFINITION BP267918 Sugano cDNA library, thyroid JTH Homo sapiens cDNA clone
JTH08419, mRNA sequence.
ACCESSION BP267918
VERSION BP267918.1 GI:52183150
KEYWORDS EST.
SOURCE Homo sapiens (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1 (bases 1 to 240)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
10737800
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2=PM4-BN0172-010
600-006-h11&t3=2000-06-01&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 11
High quality sequence stop: 79.
Location/Qualifiers
1. 240
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="BN0172"
/notes="Organ: breast normal; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

```

```

BE815636/c
LOCUS
DEFINITION PM4-BN0172-010600-006-h11 BN0172 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE815636
VERSION BE815636.1 GI:10247870
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1 (bases 1 to 240)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
10737800
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2=PM4-BN0172-010
600-006-h11&t3=2000-06-01&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 11
High quality sequence stop: 79.
Location/Qualifiers
1. 240
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="BN0172"
/notes="Organ: breast normal; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

```

```

FEATURES
source
1. 240
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="BN0172"
/notes="Organ: breast normal; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

```

```

ORIGIN
Query Match 62.4%; Score 26.2; DB 7; Length 240;
Best Local Similarity 90.3%; Pred. No. 5.9;
Matches 28; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 10 TGCCCGAGCCCTTACAAAGACTATGAAGTAA 40
| | | | | | | | | | | | | | | | | | | | |
Db 151 TTGCCGAGCCCTTACAAAGACTATGAATAAA 121
| | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 48
AA222951
LOCUS
DEFINITION mw01h08.r1 GuayWoodford Beier mouse kidney day 7 Mus musculus cDNA
clone IMAGE:663327 5', mRNA sequence.
ACCESSION AA222951
VERSION AA222951.1 GI:1843213
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1 (bases 1 to 336)
GuayWoodford Beier, M.
GuayWoodford Beier mouse kidney day 7 Mus musculus cDNA
clone IMAGE:663327 5', mRNA sequence.

```

Washington University School of Medicine

(<http://www.ludwig.org.br/scripts/gethtml2.pl?i=&t2=QV4-BN0090-040>)  
500-199-c03&t3=2000-05-04&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 13  
High quality sequence stop: 365.

## FEATURES

```
1. 374
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="BN0090"
/note="Organ: breast normal tissue; Site 2: SMAI; A mini-lib derived from ORESTES PCRs. Profiles into the pUC 119 tissue mRNA and cDNA amplification. Low stringency condition."
No. 196,716 - Ludwig In
```

## ORIGIN

```
Query Match          55.2%; Score 23.2; DB 7; Length 374;
Best Local Similarity 89.3%; Pred. No. 1.2e+02;
Matches 25; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

**Qy**

15 GAGCCCTTACAAAGACTATGAAGTAAAG 42  
| | | | | | | | | | | | | | |  
**Db**

35 GCGAGCTTACAAAGACTATGAAGTAAAG 62

RESULT 51  
CNS03FKK/0

LOCUS	CNS03FKK	907 bp	DNA	linear	GSS 01-SEP-2000
DEFINITION	Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone 02012 of library G from Tetraodon nigroviridis, genomic survey sequence.				

sequence.  
ACCESSION AL241805  
VERSION AL241805.1 GI:7962574  
KEYWORDS GSS; genome survey seq  
SOURCE Tetraodon nigroviridis  
ORGANISM Tetraodon nigroviridis

SOURCE	ORGANISM
Tetraodon nigroviridis	Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodontidae; Tetraodon.	

1  
REFERENCE  
AUTHORS  
Roest Croollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,  
Bernot, A., Fzames, C., Winkler, P., Brottier, P., Quetier, F.,  
Saurin, W. and Weissenbach, J.  
TITLE  
Estimate of human gene number provided by genome-wide analysis  
using Tetradon nigroviridis DNA sequence  
JOURNAL  
Nat. Genet. 25 (2) 235-238 (2000)

JOURNAL  
PUBMED  
REFERENCE  
Natl. Genet.  
10835645  
2

2  
REFERENCE  
AUTHORS  
Roest Crolius, H., Jaillon, O., Dasilva, C., Ozouf-Costaz, C.,  
Fizames, C., Fischer, C., Bouneau, L., Billault, A., Quetier, F.,  
Saurin, W., Bernot, A. and Weissenbach, J.  
TITLE  
Characterization and repeat analysis of the compact genome of the  
freshwater puffball Tetraodon nigroviridis  
JOURNAL  
Genome Res. 10 (7), 939-949 (2000)

JOURNAL  
PUBMED  
REFERENCE  
AUTHORS  
Genome Res. 10 (1998)  
10899143  
3 (bases 1 to 907)  
Genoscope.

**AUTHORS**  
**TITLE**  
**JOURNAL**

Genoscope.  
Direct Submission  
Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :  
Submitted (12-APR-2000) Genoscope - FRANCE (E-mail : [segref@genoscope.cns.fr](mailto:segref@genoscope.cns.fr))  
BP 191 91006 EVRY cedex - FRANCE (E-mail : [segref@genoscope.cns.fr](mailto:segref@genoscope.cns.fr))  
Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)

COMMENT  
This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.

## FEATURES source

Location/Qualifiers  
1. .907

```
/organism="Tetraodon nigroviridis"  
/mol_type="genomic DNA"  
/db_xref="taxon:99883"  
/clone="022012"  
/clone_lib="H"  
/note="Genoscope sequence ID : COBG022BH06SP1  
end : PUC-Orig"
```

## ORIGIN

```
Query Match      55.2%; Score 23.2; DB 14; Length 907;
Best Local Similarity 77.8%; Pred. No. 1.4e+02;
Matches 28; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
```

1 GCCCACGCTTGGCCGAGCCCTTACAAAGACTATGAA 36  
517 GCCAACCCCTGGCAGAACCCCTTACACAGAAATTTGAA 482

## RESULT 52

LOCUS	BP267556	578 bp	linear	EST 16-SEP-2004
DEFINITION	BP267556	578 bp	mRNA	
	BP267556	Sugano cDNA library, thyroid JTH	Homo sapiens	cDNA clone
	JTH07367			mRNA sequence.

U01073587; IMAGE sequence

ACCESSION	BP267556
VERSION	BP267556.1
KEYWORDS	EST.

GI:52182788

KEYWORDS  
EST.  
Homo sapiens (human)  
SOURCE

SOURCE: Homo sapiens (human);  
Homo sapiens  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.

REFERENCE 1 (bases 1 to 578)

**REFERENCE**  
Suzuki, K., Yamashita, R., Shirota, M., Satakiyama, Y., Chiba, J., Mizushima-Sugano, J., Nakai, K. and Sugano, S. (2004) Sequence comparison of human and mouse genes reveals a homologous block structure in the promoter regions of the genome. *Genome Res.* 14 (9), 1711-1718 (2004).

JOURNAL OF GENOME K  
PIUMED 15342556

15342556  
Contact: Yutaka Suzuki  
Department of Virology  
Institute of Medical Science, University of Tokyo  
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
Tel: 81-3-5449-5343  
Fax: 81-3-5449-5416  
Email: vsuzuki@bac.jp.

FEATURES

Email: ysuzukr@ngc.jp

Location/Qualifiers

```

FEATURES
SOURCE
LOCATION/Qualifier
1..578
/organism="Homo sapiens"
/mol_type="mrna"
/db_xref="taxon:9606"
/clone="JTH07367"
/tissue_type="thyroid"
/cell_line="JTH"
/clone_lib="Sugano cDNA library, thyroid JTH"
name="thyroid tumor"

```

ORIGINE

Query Match 54.8%; Score 23; DB 3; Length 578;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCCGAGCCCTTA 23  
db 556 GCCCAGCTTGGCCGAGCCCTTA 576

## RESULT 53

[illegible]



```

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Caenorhabditis elegans
Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 360)
Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
Sano,M., Miyata,A. and Nishigaki,A.
Expression map of the C.elegans genome
Unpublished (1996)
Contact: Yuji Kohara
Genome Biology Lab.
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: yokohara@lab.nig.ac.jp.
Location/Qualifiers
1. 360
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="N2"
/db_xref="taxon:6239"
/clone="yk394a11"
/sex="hermaphrodite"
/dev_stage="embryo"
/clone_lib="Yuji Kohara unpublished cDNA-Strain N2
hermaphrodite embryo"

ORIGIN
Query Match 54.3%; Score 22.8; DB 4; Length 360;
Best Local Similarity 71.4%; Pred. No. 1.8e+02;
Matches 30; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGGCGAGCCCTTACAAGACTATGAAGTAAAG 42
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 230 GCAGATGCTTTCGAGCCATTGGAAGACGCTGAAGGATAG 189

RESULT 54
CE847883 630 bp DNA linear GSS 01-OCT-2003
LOCUS tigr-gss-dog-17000332951017 Dog Library Canis familiaris genomic,
genomic survey sequence.
CE847883
VERSION CE847883.1 GI:37212952
KEYWORDS GSS.
SOURCE Canis familiaris (dog)
ORGANISM Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
Canis.
1 (bases 1 to 630)
Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
Ruech,D.B., Delcher,A.L., Pop.M., Wang,W., Fraser,C.M. and
Venter,J.C.
The dog genome: survey sequencing and comparative analysis
Science 301 (5641), 1898-1903 (2003)
14512627
Contact: Kirkness EF
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
Rockville, MD 20850, USA
Tel: 301-838-0200
Fax: 301-838-0208
Email: ekirknes@tigr.org
Class: shotgun.
Location/Qualifiers
1. 630
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="Standard Poodle"
/db_xref="taxon:9615"

FEATURES
source
Query Match 54.3%; Score 22.8; DB 8; Length 715;
Best Local Similarity 71.4%; Pred. No. 2e+02;
Matches 30; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGGCGAGCCCTTACAAGACTATGAAGTAAAG 42
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 218 GTCCAGACTCGGGTTAGCCCTTACAGTGACTACGAAGGAAGG 259

RESULT 55
CK976767 715 bp mRNA linear EST 16-MAR-2004
LOCUS 4107795 BARC 9BOV Bos taurus cDNA clone 9BOV34_E02 5', mRNA
sequence.
ACCESSION CK976767
VERSION CK976767.1 GI:45494741
KEYWORDS EST.
SOURCE Bos taurus (cattle)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
Pecora; Bovidae; Bovinae; Bos.
1 (bases 1 to 715)
Sonstegard, T.S., Van Tassel, C.P., Matukumalli, L.K., Harhay,
G.P., Bosak, S., Rubenfield, M. and Gasbarre, L.C.
Production of EST from cDNA libraries derived from immunologically
activated bovine gut
Unpublished (2004)
Contact: Tad S. Sonstegard
Bovine Functional Genomics Laboratory
Animal and Natural Resources Institute
Bldg. 200 Rm2A BARC-East, Beltsville, MD 20705, USA
Tel: 3015048416
Fax: 3015048414
Email: tads@nri.barc.usda.gov
Single pass sequencing. Bases called and trimmed with phred
0.000925 using options -trim alt " -trim fasta. Vector identified
by cross match using options -minmatch 12 -minscore 12
Plate: 34 row: E column: 02
Seq primer: CCCAGTCACGACGTTGTAACG
High quality sequence stop: 715.
Location/Qualifiers
1. 715
/organism="Bos taurus"
/mol_type="mRNA"
/strain="Holstein"
/db_xref="taxon:9913"
/clone="9BOV34_E02"
/sex="Male"
/tissue_type="Pooled"
/dev_stage="Multiple"
/lab_host="DH10B T1 phage resistant"
/clone_lib="BARC 9BOV"
/notes="Organ: Abomasum; Vector: pAgen-1; Site_1: EcoRV;
Site_2: NotI; Equimolar amounts of mRNA extracted from
fundic and pyloric abomasums of 18 and 21 week old steers.
Exposure to Osteragia osteragi was initiated at 15 weeks
of age. fundic and pyloric abomasum"

ORIGIN
Query Match 54.3%; Score 22.8; DB 8; Length 715;
Best Local Similarity 71.4%; Pred. No. 2e+02;
Matches 30; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGGCGAGCCCTTACAAGACTATGAAGTAAAG 42
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 218 GTCCAGACTCGGGTTAGCCCTTACAGTGACTACGAAGGAAGG 259

```

```

/clone_lib="Dog Library"
/notes="Site 1: BstXI; Libraries were prepared from
peripheral blood"

ORIGIN
Query Match 54.3%; Score 22.8; DB 12; Length 630;
Best Local Similarity 71.4%; Pred. No. 2e+02;
Matches 30; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGGCGAGCCCTTACAAGACTATGAAGTAAAG 42
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 483 GCCTGCGCTTGGACTAGCCCTGACCTAGACTCTCAAAACCAAG 524

RESULT 55
CK976767 715 bp mRNA linear EST 16-MAR-2004
LOCUS 4107795 BARC 9BOV Bos taurus cDNA clone 9BOV34_E02 5', mRNA
sequence.
ACCESSION CK976767
VERSION CK976767.1 GI:45494741
KEYWORDS EST.
SOURCE Bos taurus (cattle)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
Pecora; Bovidae; Bovinae; Bos.
1 (bases 1 to 715)
Sonstegard, T.S., Van Tassel, C.P., Matukumalli, L.K., Harhay,
G.P., Bosak, S., Rubenfield, M. and Gasbarre, L.C.
Production of EST from cDNA libraries derived from immunologically
activated bovine gut
Unpublished (2004)
Contact: Tad S. Sonstegard
Bovine Functional Genomics Laboratory
Animal and Natural Resources Institute
Bldg. 200 Rm2A BARC-East, Beltsville, MD 20705, USA
Tel: 3015048416
Fax: 3015048414
Email: tads@nri.barc.usda.gov
Single pass sequencing. Bases called and trimmed with phred
0.000925 using options -trim alt " -trim fasta. Vector identified
by cross match using options -minmatch 12 -minscore 12
Plate: 34 row: E column: 02
Seq primer: CCCAGTCACGACGTTGTAACG
High quality sequence stop: 715.
Location/Qualifiers
1. 715
/organism="Bos taurus"
/mol_type="mRNA"
/strain="Holstein"
/db_xref="taxon:9913"
/clone="9BOV34_E02"
/sex="Male"
/tissue_type="Pooled"
/dev_stage="Multiple"
/lab_host="DH10B T1 phage resistant"
/clone_lib="BARC 9BOV"
/notes="Organ: Abomasum; Vector: pAgen-1; Site_1: EcoRV;
Site_2: NotI; Equimolar amounts of mRNA extracted from
fundic and pyloric abomasums of 18 and 21 week old steers.
Exposure to Osteragia osteragi was initiated at 15 weeks
of age. fundic and pyloric abomasum"

ORIGIN
Query Match 54.3%; Score 22.8; DB 8; Length 715;
Best Local Similarity 71.4%; Pred. No. 2e+02;
Matches 30; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGGCGAGCCCTTACAAGACTATGAAGTAAAG 42
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 218 GTCCAGACTCGGGTTAGCCCTTACAGTGACTACGAAGGAAGG 259

```

```

RESULT 56
DN020621
LOCUS
DEFINITION
  JGI CAAR2620.rev NTH XGC tropLiv1 Xenopus tropicalis cDNA clone
  IMAGE:7734970 3', mRNA sequence.
ACCESSION
DN020621
VERSION
KEYWORDS
SOURCE
  DN020621.1 GI:58864255
  EST.
  Xenopus tropicalis (western clawed frog)
  ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
    Xenopodinae; Xenopus; Silurana.
  1 (bases 1 to 864)
  Richardson,P., Lucas,S., Rokhsar,D., Dettter,J.C., Ng,D.C.,
  Brokstein,P. and Lindquist,E.A.
  DOE Joint Genome Institute Xenopus tropicalis EST project
  Unpublished (2004)
  Other ESTs: JGI CAAR2620.fwd
  Contact: Lindquist,E.A., Richardson,P.
  DOE Joint Genome Institute
  2800 Mitchell Drive, Walnut Creek, CA 94598, USA
  Tel: 925 296 5600
  Fax: 925 296 5710
  Email: cdna@jgi-psf.org
  Tissue Procurement: Robert M. Grainger
  cDNA Library Preparation: Bruce Blumberg Laboratory, University of
  California, Irvine
  DNA Sequencing: DOE Joint Genome Institute: http://www.jgi.doe.gov
  Clone Distribution: I.M.A.G.E. Consortium/LLNL:
  http://image.llnl.gov
  Naming Conventions: EST name is generated by the concatenation of
  the JGI Clone id and the direction of sequencing. The suffix '.rev'
  indicates a reverse sequencing read of the insert. It does not
  necessarily reflect the orientation of the insert.
  Poly-A: Based upon the presence of a run of 14 or more T residues
  at the beginning of the sequence, this clone was polyadenylated.
  The resulting Poly-T sequence has been removed.
  Plate: CAAR 0025 row: h column: 8
  High quality sequence stop: 751
  POLYA=Yes.
  Location/Qualifiers
    1..864
      /organism="Xenopus tropicalis"
      /mol_type="mRNA"
      /strain="N6 (Nigerian 6th generation inbred)"
      /db_xref="taxon:8164"
      /clone="IMAGE:7734970"
      /sex="male"
      /tissue_type="Liver"
      /dev_stage="Adult"
      /lab_host="NTH XGC tropLiv1"
      /clone_lib="NTH XGC tropLiv1"
      /note="Vector: pCS107; Site_1: EcoRI; Site_2: XhoI; The
      library was prepared from 5 ug of poly A+ RNA by oligo-dT
      priming
      (5'-GAGAGAGAGAGAGAGACTAGTCTCGAGTGTCTTTT-3')
      and Stratascript reverse transcriptase. After ligation of
      EcoRI adapters (5'-AATTCGACGACGAGG-3') followed by kinasing
      adapters and by XhoI digestion, the cDNA was size selected
      by chromatography on Sepharose CL-2B columns and fractions
      containing cDNAs larger than 1000 bp were ligated into
      EcoRI/XhoI-digested pCS107. Reference for library
      construction: Current Genomics 4, 635-644. Library
      constructed by Michelle Tabb and Bruce Blumberg (Dept of
      Developmental and Cell Biology, University of California,
      Irvine)."
  Query Match 54.3%; Score 22.8; DB 9; Length 864;
  Best Local Similarity 77.1%; Pred. No. 2.1e+02;
  Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

RESULT 57
AA615518
LOCUS
DEFINITION
  v070e02.r1 Barstead mouse myotubes MPLRB5 Mus musculus cDNA clone
  IMAGE:1064474 5', mRNA sequence.
ACCESSION
AA615518
VERSION
KEYWORDS
SOURCE
  AA615518.1 GI:2502746
  EST.
  Mus musculus (house mouse)
  ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
    Sciurognathi; Muroidae; Muridae; Murinae; Mus.
  1 (bases 1 to 454)
  Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
  Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
  Scheilenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
  Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
  Waterston,R.
  The WashU-HMNI Mouse EST Project
  Unpublished (1996)
  Contact: Marra M/Mouse EST Project
  WashU-HMNI Mouse EST Project
  Washington University School of Medicine
  4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
  Tel: 314 286 1800
  Fax: 314 286 1810
  Email: mouseest@watson.wustl.edu
  This clone is available royalty-free through LLNL ; contact the
  IMAGE Consortium (info@image.llnl.gov) for further information.
  MGI:586834
  Seq primer: -28ml3 rev2 ET from Amersham.
  Location/Qualifiers
    1..454
      /organism="Mus musculus"
      /mol_type="mRNA"
      /strain="C3H"
      /db_xref="taxon:10090"
      /clone="IMAGE:1064474"
      /cell_line="C2C12"
      /lab_host="DH10B"
      /clone_lib="Barstead mouse myotubes MPLRB5"
      /note="Vector: pT7T3D-PacI; Site 1: EcoRI; Site 2: NotI;
      1st strand cDNA was primed with a Not I - oligo(dT) primer
      [5.
      TGTTACGAATCTGAAGTGGAGCGCGCCCTTTT-3'
      3']; double-stranded cDNA was ligated to Eco RI adaptors
      [AATTCGATCCTTG], digested with Not I and cloned into the
      Not I and Eco RI sites of the modified pT7T3 vector.
      Library constructed by Bob Barstead. The C2C12 cell line
      (available from ATCC, catalog # CRL-1772) differentiates
      rapidly, forming contractile myotubes and producing
      characteristic muscle proteins. "
  Query Match 53.8%; Score 22.6; DB 1; Length 454;
  Best Local Similarity 75.7%; Pred. No. 2.3e+02;
  Matches 28; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

RESULT 58
CB818826
LOCUS
DEFINITION
  1 GCCCAGCGTGGCCGAGCCCTTACAAAGACTATGAAG 37
  326 GCCCAGCGCCGCCCATGTCCGTACATGACTACGAAG 362
  Query Match 54.3%; Score 22.8; DB 9; Length 864;
  Best Local Similarity 77.1%; Pred. No. 2.1e+02;
  Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```

```

Qy 8 CTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42
  ||||| ||| ||||| ||| |||||
Db 820 CTTGGACGACATTTACAAAGCATATNCAGTAAAG 854

RESULT 57
AA615518
LOCUS
DEFINITION
  v070e02.r1 Barstead mouse myotubes MPLRB5 Mus musculus cDNA clone
  IMAGE:1064474 5', mRNA sequence.
ACCESSION
AA615518
VERSION
KEYWORDS
SOURCE
  AA615518.1 GI:2502746
  EST.
  Mus musculus (house mouse)
  ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
    Sciurognathi; Muroidae; Muridae; Murinae; Mus.
  1 (bases 1 to 454)
  Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
  Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
  Scheilenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
  Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
  Waterston,R.
  The WashU-HMNI Mouse EST Project
  Unpublished (1996)
  Contact: Marra M/Mouse EST Project
  WashU-HMNI Mouse EST Project
  Washington University School of Medicine
  4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
  Tel: 314 286 1800
  Fax: 314 286 1810
  Email: mouseest@watson.wustl.edu
  This clone is available royalty-free through LLNL ; contact the
  IMAGE Consortium (info@image.llnl.gov) for further information.
  MGI:586834
  Seq primer: -28ml3 rev2 ET from Amersham.
  Location/Qualifiers
    1..454
      /organism="Mus musculus"
      /mol_type="mRNA"
      /strain="C3H"
      /db_xref="taxon:10090"
      /clone="IMAGE:1064474"
      /cell_line="C2C12"
      /lab_host="DH10B"
      /clone_lib="Barstead mouse myotubes MPLRB5"
      /note="Vector: pT7T3D-PacI; Site 1: EcoRI; Site 2: NotI;
      1st strand cDNA was primed with a Not I - oligo(dT) primer
      [5.
      TGTTACGAATCTGAAGTGGAGCGCGCCCTTTT-3'
      3']; double-stranded cDNA was ligated to Eco RI adaptors
      [AATTCGATCCTTG], digested with Not I and cloned into the
      Not I and Eco RI sites of the modified pT7T3 vector.
      Library constructed by Bob Barstead. The C2C12 cell line
      (available from ATCC, catalog # CRL-1772) differentiates
      rapidly, forming contractile myotubes and producing
      characteristic muscle proteins. "
  Query Match 53.8%; Score 22.6; DB 1; Length 454;
  Best Local Similarity 75.7%; Pred. No. 2.3e+02;
  Matches 28; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

RESULT 58
CB818826
LOCUS
DEFINITION
  1 GCCCAGCGTGGCCGAGCCCTTACAAAGACTATGAAG 37
  326 GCCCAGCGCCGCCCATGTCCGTACATGACTACGAAG 362
  Query Match 54.3%; Score 22.8; DB 9; Length 864;
  Best Local Similarity 77.1%; Pred. No. 2.1e+02;
  Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```

EST 1716 Green Apricot Fruit Lambda Zap II Library Prunus armeniaca  
cDNA clone ba003d02 3', mRNA sequence.

CB818826  
CB818826.1 GI:40715236  
EST.

Prunus armeniaca (apricot)  
Prunus armeniaca

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
rosids; eurosids I; Rosales; Rosaceae; Amygdaloideae; Prunus.  
1 (bases 1 to 754)

Bouquet, J., Romieu, C., Audergon, J.M., Albagnac, G., Lambert, P.,  
Bouchet, J.P. and Terrier, N.

High Throughput Detection of Isogenes among 5724 3' EST from  
Apricot Fruit (Prunus armeniaca)

Unpublished (2003)  
Contact: Audergon JM

Unité de génétique et amélioration des fruits et légumes  
Institut National de la Recherche Agronomique

Domaine Saint-Maurice BP 94 84143 Montfavet cedex  
Tel: 00-33-(0)4-32-72-26-68

Fax: 00-33-(0)4-32-72-26-62  
Email: audergon@avignon.inra.fr

Seq primer: 17.  
Location/Qualifiers

1..754  
/organism="Prunus armeniaca"  
/mol\_type="mRNA"  
/cultivar="Bergeron"  
/db\_xref="taxon:36596"  
/clone="ba003d02"  
/dev\_stage="Green stage"  
/clone\_lib="Green Apricot Fruit Lambda Zap II Library"  
/note="Organ: Fruit; Vector: Lambda Zap II; Site\_1: Eco  
RI; Site\_2: XhoI; Oriented library, construction described  
in Molecular cloning and expression of a cDNA encoding  
1-aminocyclopropane-1-carboxylate (ACC) oxidase from  
apricot fruit (Prunus armeniaca cv. Bergeron) by  
Mbequie-Mbequie D, Chahine H, Gomez RM, Gouble B, Audergon  
JM, Soury W, Albagnac G, Flis-Lyaon B in Physiol Plant  
105:294-303 1999"

FEATURES  
source

Query Match 53.8%; Score 22.6; DB 4; Length 754;  
Best Local Similarity 75.7%; Pred. No. 2.5e+02;  
Matches 28; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

OY 4 CACGCTGGCCGAGCCCTTACAAAGACTATGAAGTAA 40  
|||||  
Db 89 CACCATGGCGCGTCCGTACAAAGACCGTGAAGAAA 125  
|||||

ORIGIN

RESULT 59  
DV624038  
LOCUS  
DEFINITION  
93529.1 Cold Sweetening C Solanum tuberosum cDNA clone 93529 5',  
mRNA sequence.

DV624038  
DVG24038.1 GI:78747635  
EST.

Solanum tuberosum (potato)  
Solanum tuberosum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
asterids; lamiales; Solanales; Solanaceae; Solanum.  
1 (bases 1 to 609)

Flinn, B., Rothwell, C., Sardana, R., Griffiths, R., Lague, M., De  
Koeyer, D., Audy, P., Gover, C., Li, X.-Q., Wang-Pruski, G. and Regan, S.

Generation of ESTs from tubers following 3 months storage at 4  
degrees Celsius, and 95% relative humidity, then 3 days of  
reconditioning at 21 degrees

Unpublished (2005)  
Contact: Vicki Gustafson

JOURNAL  
COMMENT

The Canadian Potato Genome Project - BioAtlantech  
921 College Hill Rd, Fredericton, ON, E3B 6Z9, CANADA  
Email: ygustafson@bioatlantech.nb.ca  
Seq primer: 13.

FEATURES  
source

Location/Qualifiers  
1..609  
/organism="Solanum tuberosum"  
/mol\_type="mRNA"  
/cultivar="Shepody"  
/db\_xref="taxon:4113"  
/clone="93529"  
/tissue\_type="Tubers"  
/lab\_host="XL10-Gold"  
/clone\_lib="Cold Sweetening C"  
/note="Vector: pBluescript II SK(+) XR; Site\_1: EcoRI;  
Site\_2: XhoI; supplier: Developmental series. Plants from  
pathogen-free Solanum tuberosum var. Shepody, clone 1756,  
nuclear stock were grown in a greenhouse under natural  
conditions. Mature, harvested tubers were stored in the  
dark at 4C, 95% relative humidity for 3 months, and then  
reconditioned at 21C for three days. RNA was isolated for  
library construction."

ORIGIN

Query Match 53.3%; Score 22.4; DB 10; Length 609;  
Best Local Similarity 72.5%; Pred. No. 2.9e+02;  
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

OY 2 CCCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAA 41  
|||||  
Db 149 CCATTGCTTGCACCAAGCCATGAACAAGATTATGAAGTCAA 188  
|||||

RESULT 60  
AQ855455/c

LOCUS  
DEFINITION

CPG1823A CpIOWAGDNL Cryptosporidium parvum genomic, genomic survey  
sequence.

ACCESSION  
AQ855455

VERSION  
AQ855455.1 GI:6205912

KEYWORDS  
GSS.

SOURCE  
Cryptosporidium parvum

ORGANISM  
Cryptosporidium parvum

REFERENCE  
AUTHORS

TITLE  
PUBLISHED

COMMENT  
Strong, W.B. and Nelson, R.G.

Preliminary profile of the Cryptosporidium parvum genome: an  
expressed sequence tag and genome survey sequence analysis

Mol. Biochem. Parasitol. 107 (1), 1-32 (2000)  
10717299

Contact: Nelson, R. G.  
Depts. of Medicine & Pharmaceutical Chemistry

San Francisco General Hospital-University of California, San  
Francisco

Box 0811, San Francisco, CA 94143-0811, USA  
Tel: 415 206 8846

Fax: 415 206 3353  
Email: malaria@itsa.ucsf.edu

For Annotation Data see http://medsfgh.ucsf.edu/id/CpTags/home.html  
Seq primer: M13(-21) forward

Class: shotgun.  
Location/Qualifiers

1..746  
/organism="Cryptosporidium parvum"  
/mol\_type="genomic DNA"  
/strain="IOWA"  
/db\_xref="taxon:5807"  
/lab\_host="E. coli XL2 Blue MRF"  
/clone\_lib="CpIOWAGDNL"  
/note="Vector: pBluescript II (SK-); Site 1: EcoRV; C.  
parvum (IOWA isolate) genomic DNA was hydrodynamically  
sheared to produce fragments having a tight size



TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

PUBMED 10737800

COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?l1=6t2=CMO-BN0105-180  
300-295-e05&t3=2000-03-18&t4=1)

Seq primer: puc 18 forward  
High quality sequence stop: 336.

## FEATURES

source

```

1..338
    Location/Qualifiers
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /dev_stage="Adult"
        /clone_lib="BN0105"
        /note="Organ: breast normal; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

```

## ORIGIN

```

Query Match      52.9%; Score 22.2; DB 7; Length 338;
Best Local Similarity 77.1%; Pred. No. 3.2e+02;
Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 7 GCTTGGCGGAGCCCTTACAAAGACTATGAAGTAAA 41
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 231 GGTGGCCCAACTCTCACAAATACAAAGAGTAAA 265.

```

```

RESULT 64
AA956676
LOCUS      362 bp      mRNA      linear      EST 04-JUL-1999
DEFINITION      UI-R-EI-fl-g-03-0-UI.s1 UI-R-EI Rattus norvegicus cDNA clone
                UI-R-EI-fl-g-03-0-UI 3', similar to gi|2888286|gb|AA818700|AA818700
                UI-R-A0-az-c-06-0-UI.s1 UI-R-A0 Rattus norvegicus cDNA clone
                UI-R-A0-az-c-06-0-UI 3', similar to gi|1808536|gb|AA210199|AA210199
                mu40c01.r1 Soares 2NbMT Mus musculus cDNA clone 641856 5', mRNA
                sequence.

```

```

ACCESSION      AA956676
VERSION        AA956676.1 GI:4239150
KEYWORDS
SOURCE
ORGANISM      Rattus norvegicus (Norway rat)
                Rattus norvegicus
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
                Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
                1 (bases 1 to 362)
                Bonaldo,M.F., Lennon,G. and Soares,M.B.
                Normalization and subtraction: two approaches to facilitate gene
                discovery
                Genome Res. 6 (9), 791-806 (1996)
                8889548
                On May 7, 1998 this sequence version replaced gi:3120371.
                Contact: Soares, MB
                Coordinated Laboratory for Computational Genomics
                University of Iowa
                375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA
                Tel: 319 335 8250
                Fax: 319 335 9565

```

```

REFERENCE
AUTHORS
TITLE
JOURNAL
PUBMED
COMMENT

```

## FEATURES

source

```

1..362
    Location/Qualifiers
        /organism="Rattus norvegicus"
        /mol_type="mRNA"
        /strain="Sprague-Dawley"
        /db_xref="taxon:10116"
        /clones="UI-R-EI-fl-g-03-0-UI"
        /dev_stage="adult"
        /lab_host="DH10B (Life Technologies)"
        /clone_lib="UI-R-EI"
        /note="Vector: p773D-PacI; Site 1: Not I; Site 2: Eco RI; The UI-R-EI library is a subtracted library derived from the UI-R-E0 library. The UI-R-E0 library consisted of a mixture of individually tagged normalized libraries constructed from 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-EI) was constructed as follows: PCR amplified cDNA inserts from a pool of UI-R-E0 clones from which 3' ESTs had been derived was used as a driver in a hybridization with the UI-R-E0 library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-EI library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996)."

```

## ORIGIN

```

Query Match      52.9%; Score 22.2; DB 1; Length 362;
Best Local Similarity 88.9%; Pred. No. 3.2e+02;
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 15 GAGCCCTTACAAAGACTATGAAGTAAA 41
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 270 GAGCCCTTACAAAGACTATGAAGTAAA 296

```

```

RESULT 65
CB727988
LOCUS      482 bp      mRNA      linear      EST 11-APR-2003
DEFINITION      AMGNNUC:URRG1-00029-C2-A urrg1 (14046) Rattus norvegicus cDNA clone
                urrg1-00029-c2 5', mRNA sequence.

```

```

ACCESSION      CB727988
VERSION        CB727988.1 GI:29795061
KEYWORDS
SOURCE
ORGANISM      Rattus norvegicus (Norway rat)
                Rattus norvegicus
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
                Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
                1 (bases 1 to 482)
                Angen EST Program.
                Angen Rat EST Program
                Unpublished (2003)
                Contact: Dan Fitzpatrick
                Angen, Inc
                One Angen Center Drive, Thousand Oaks, CA 91320-1799, USA
                Tel: 805 447-4881

```

```

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```

```

FEATURES
source
Plate: 00029 row: c column: 2.
Location/Qualifiers
1. .482
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="urrgi-00029-c2"
/clone_lib="urrgi (14046)"
/notes="Vector: pT7T3D-Paci"

ORIGIN
Query Match 52.9%; Score 22.2; DB 4; Length 482;
Best Local Similarity 88.9%; Pred. No. 3.4e+02;
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 15 GAGCCCTTACAAAGACTATGAAGTAA 41
|||||
Db 403 GAGCCCTTACAAAGACTTAAAGACCAA 429
|||||

RESULT 66
A1171672
LOCUS
DEFINITION
EST217647 Normalized rat muscle, Bento Soares Rattus sp. cDNA clone
RMUB014.3, end, mRNA sequence.
ACCESSION
A1171672
VERSION
A1171672.1 GI:3711712
KEYWORDS
EST.
SOURCE
Rattus sp.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE
1 (bases 1 to 484)
Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
Kerlavage,A.R. and Adams,M.D.
Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
Gene Index
Unpublished (1998)
Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.
Site_2: NotI"

FEATURES
source
Location/Qualifiers
1. .484
/organism="Rattus sp."
/mol_type="mRNA"
/db_xref="ATCC (inhost):2028324"
/clone="RMUB014"
/clone_lib="Normalized rat muscle, Bento Soares"
/notes="Organ: muscle; Vector: pT7T3Pac; Site_1: EcoRI;
Site_2: NotI"

ORIGIN
Query Match 52.9%; Score 22.2; DB 1; Length 484;
Best Local Similarity 88.9%; Pred. No. 3.4e+02;
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 15 GAGCCCTTACAAAGACTATGAAGTAA 41
|||||
Db 356 GAGCCCTTACAAAGACTTAAAGACCAA 382
|||||

RESULT 67
BP470636/c
LOCUS
DEFINITION
BP470636 Rattus norvegicus pancreatic islet Rattus norvegicus cDNA
clone RBC06466 3', mRNA sequence.
ACCESSION
BP470636

```

```

VERSION
KEYWORDS
SOURCE
ORGANISM
BP470636.1 GI:85732193
EST.
Rattus norvegicus (Norway rat)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE
1 (bases 1 to 498)
Takeda,J., Wang,H. and Horikawa,Y.
Expression profile of mRNAs from rat pancreatic islet
Unpublished (2006)
Contact: Yukio Horikawa
Laboratory of Molecular Genetics
Institute for Molecular and Cellular Regulation, Gunma University
Showa-machi 3-39-15, Maebashi, Gunma, 371-8512, Japan
Tel: 81-27-220-8832
Fax: 81-27-220-8889
Email: yhorikaw@showa.gunma-u.ac.jp.

FEATURES
source
Location/Qualifiers
1. .498
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="RBC06466"
/tissue_type="pancreatic islet"
/clone_lib="Rattus norvegicus pancreatic islet"

ORIGIN
Query Match 52.9%; Score 22.2; DB 3; Length 498;
Best Local Similarity 88.9%; Pred. No. 3.4e+02;
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 15 GAGCCCTTACAAAGACTATGAAGTAA 41
|||||
Db 420 GAGCCCTTACAAAGACTTAAAGACCAA 394
|||||

RESULT 68
A1175079
LOCUS
DEFINITION
EST218597 Normalized rat muscle, Bento Soares Rattus sp. cDNA clone
RMUCCL15.3, end, mRNA sequence.
ACCESSION
A1175079
VERSION
A1175079.1 GI:3725717
KEYWORDS
EST.
SOURCE
Rattus sp.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE
1 (bases 1 to 515)
Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
Kerlavage,A.R. and Adams,M.D.
Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
Gene Index
Unpublished (1998)
Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.
Location/Qualifiers
1. .515
/organism="Rattus sp."
/mol_type="mRNA"
/db_xref="ATCC (inhost):2032225"
/clone="RMUCCL15"
/clone_lib="Normalized rat muscle, Bento Soares"
/notes="Organ: muscle; Vector: pT7T3Pac; Site_1: EcoRI;
Site_2: NotI"

FEATURES
source
Location/Qualifiers
1. .515
/organism="Rattus sp."
/mol_type="mRNA"
/db_xref="ATCC (inhost):2032225"
/clone="RMUCCL15"
/clone_lib="Normalized rat muscle, Bento Soares"
/notes="Organ: muscle; Vector: pT7T3Pac; Site_1: EcoRI;
Site_2: NotI"

```

## ORIGIN

Query Match 52.9%; Score 22.2; DB 1; Length 515;  
 Best Local Similarity 88.9%; Pred. No. 3.4e+02;  
 Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 15 GAGCCCTTACAAAGACTATGAAGTAAA 41  
 |||||  
 Db 390 GAGCCCTTACAAAGACTAAGAACCATA 416

## RESULT 69

AII13032

LOCUS

DEFINITION AII13032 UI-R-E1-gp-h-05-0-UI.s1 526 bp mRNA linear EST 11-FEB-1999

UI-R-E1-gp-h-05-0-UI 3', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Rattus norvegicus (Norway rat)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 526)

Bonaldo, M.F., Lennon, G. and Soares, M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

8889548

Contact: Soares, MB

Coordinated Laboratory for Computational Genomics

University of Iowa

375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: bento-soares@uiowa.edu

The sequence tag present in the cDNA between the NotI site and the

oligo-dT track served to identify it as a clone from the normalized

adult 8-day-embryo library. cDNA Library Preparation: M. Fatima

Bonaldo, Ph.D. Clone distribution: clones will be available through

Research Genetics The following repetitive elements were found in

this cDNA sequence: 1-48, &gt;AT\_rich#Low\_complexity

Seq primer: M13 Forward.

Location/Qualifiers

1..526

/organism="Rattus norvegicus"

/mol\_type="mRNA"

/strain="Sprague-Dawley"

/db\_xref="taxon:10116"

/clones="UI-R-E1-gp-h-05-0-UI"

/dev\_stage="adult"

/lab\_host="DH10B (Life Technologies)"

/clone\_lib="UI-R-E1"

/notes="Vector: pT7T3D-PacI; Site 1: Not I; Site 2: Eco RI;

The UI-R-E1 library is a subtracted library derived from

the UI-R-E0 library. The UI-R-E0 library consisted of a

mixture of individually tagged normalized libraries

constructed from 8, 12 and 18-day embryos. The tag is a

string of 3-5 nucleotides present between the Not I site

and the oligo-dT track which allows identification of the

library of origin of a clone within the mixture. The

subtracted library (UI-R-E1) was constructed as follows:

PCR amplified cDNA inserts from a pool of UI-R-E0 clones

from which 3' ESTs had been derived was used as a driver

in a hybridization with the UI-R-E0 library in the form of

single-stranded circles. The remaining single-stranded

circles (subtracted library) was purified by

hydroxyapatite column chromatography, converted to

double-stranded circles and electroporated into DH10B

bacteria (Life Technologies) to generate the UI-R-E1

library. This procedure has been previously described

(Bonaldo, Lennon and Soares, Genome Research 6: 791-806,

1996)"

Query Match 52.9%; Score 22.2; DB 1; Length 526;  
 Best Local Similarity 88.9%; Pred. No. 3.4e+02;  
 Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 15 GAGCCCTTACAAAGACTATGAAGTAAA 41  
 |||||  
 Db 385 GAGCCCTTACAAAGACTAAGAACCATA 411

## RESULT 70

BE116174

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Rattus norvegicus (Norway rat)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 532)

Bonaldo, M.F., Lennon, G. and Soares, M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

8889548

Contact: Soares, MB

Coordinated Laboratory for Computational Genomics

University of Iowa

375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: bento-soares@uiowa.edu

The sequence contained an oligo-dT track that was present in the

oligonucleotide that was used to prime the synthesis of first

strand cDNA and therefore this may represent a bonafide poly A

tail. The sequence tag present in the cDNA between the NotI site

and the oligo-dT track served to verify it as a clone from the

normalized embryo at 13 dpc library cDNA Library Preparation: M.B.

Soares Lab Clone distribution: clones will be available through

Research Genetics (www.resgen.com)

Seq primer: M13 Forward

POLYA=Yes.

Location/Qualifiers

1..532

/organism="Rattus norvegicus"

/mol\_type="mRNA"

/strain="Sprague-Dawley"

/db\_xref="taxon:10116"

/clones="UI-R-Bs1-aya-c-07-0-UI"

/dev\_stage="embryonic 13 dpc"

/lib\_host="DH10B (Life Technologies)"

/clone\_lib="UI-R-Bs1"

/notes="Vector: pT7T3D-PacI; Site 1: Not I; Site 2: Eco RI;

The UI-R-Bs1 library is derived from 13 dpc whole embryo

tissue. For a detailed description of the library from

which this clone was derived, please visit our web site

at rateat.eng.uiowa.edu.

TAG\_TISSUE=embryo at 13 dpc

TAG\_LIB=UI-R-Bs1

TAG\_SEQ=AATCC"

ORIGIN

Query Match 52.9%; Score 22.2; DB 7; Length 532;  
 Best Local Similarity 88.9%; Pred. No. 3.4e+02;  
 Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 15 GAGCCCTTACAAAGACTATGAAGTAAA 41  
 |||||

```

|||||
373 GAGCCCTTACAAAGACTAAGACCAA 399

RESULT 71
AL170330
LOCUS
DEFINITION EST216256 Normalized rat lung, Bento Soares Rattus sp. cDNA clone
RUUCH03 3' end, mRNA sequence.
ACCESSION A1170330
VERSION A1170330
KEYWORDS GI:3710370
SOURCE Rattus sp.
ORGANISM Rattus sp.
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Rattus.
REFERENCE
1 (bases 1 to 536)
Lee, N.H., Glodek, A., Chandra, I., Mason, T.M., Quackenbush, J.,
Kerlavage, A.R. and Adams, M.D.
Rat Genome Project: Generation of a Rat EST (RESTR) Catalog & Rat
Gene Index
Unpublished (1998)
Other ESTs: TCS0977
Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.
FEATURES
source
1..536
Location/Qualifiers
/organism="Rattus sp."
/mol_type="mRNA"
/db_xref="ATCC (inhost):2027795"
/db_xref="taxon:10118"
/clones="RUUCH03"
/clone_lib="Normalized rat lung, Bento Soares"
/note="Organ: lung; Vector: p7T3pac; Site_1: EcoRI;
Site_2: NotI"
ORIGIN
Query Match 52.9%; Score 22.2; DB 1; Length 536;
Best Local Similarity 88.9%; Pred. No. 3.4e+02;
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 15 GAGCCCTTACAAAGACTAAGATGAAA 41
|||||
355 GAGCCCTTACAAAGACTAAGACCAA 381

RESULT 72
CF978837/c
LOCUS
DEFINITION FAR004B12.R Rat retinal ganglion cell Rattus norvegicus cDNA, mRNA
sequence.
ACCESSION CF978837
VERSION CF978837.1
KEYWORDS GI:49174295
SOURCE EST.
Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Rattus.
REFERENCE
1 (bases 1 to 545)
Farkas, R.H., Qian, J., Goldberg, J.L., Quigley, H.A. and Zack, D.J.
Gene expression profiling of purified rat retinal ganglion cells
Invest. Ophthalmol. Vis. Sci. 45 (8), 2503-2513 (2004)
15277470
PUBMED
Contact: Farkas RH
Department of Ophthalmology
Johns Hopkins University School of Medicine

FEATURES
source
1..545
Location/Qualifiers
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/tissue_type="Retinal Ganglion Cells"
/lab_host="DH10B"
/clone_lib="Rat retinal ganglion cell"
/note="Organ: Eye; Vector: pDNK-LIB; Site_1: SfiI; Site_2:
SfiI; The library was constructed from purified rat
retinal ganglion cells. The Creator SMART cDNA Library
method (Clontech) was used. EST analysis was performed on
the unamplified, non-normalized, non-subtracted library."
ORIGIN
Query Match 52.9%; Score 22.2; DB 5; Length 545;
Best Local Similarity 88.9%; Pred. No. 3.4e+02;
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 15 GAGCCCTTACAAAGACTATGAAGTAAA 41
|||||
388 GAGCCCTTACAAAGACTAAGACCAA 362

RESULT 73
BP499326
LOCUS
DEFINITION BP499326 Rattus norvegicus insulinoma R1Nm5F Rattus norvegicus cDNA
clone R117749 3', mRNA sequence.
ACCESSION BP499326
VERSION BP499326.1
KEYWORDS GI:85983918
SOURCE EST.
Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Rattus.
REFERENCE
1 (bases 1 to 552)
Takeda, J., Wang, H. and Horikawa, Y.
Expression profile of mRNAs from Rat insulinoma cell line (RIN)
Unpublished (2006)
Contact: Yukio Horikawa
Laboratory of Molecular Genetics
Institute for Molecular and Cellular Regulation, Gunma University
Showa-machi 3-39-15, Maebashi, Gunma, 371-8512, Japan
Tel: 81-27-220-8832
Fax: 81-27-220-8889
Email: yhorikaw@showa.gunma-u.ac.jp.
FEATURES
source
1..552
Location/Qualifiers
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="R117749"
/cell_line="insulinoma R1Nm5F"
/clone_lib="Rattus norvegicus insulinoma R1Nm5F"
ORIGIN
Query Match 52.9%; Score 22.2; DB 3; Length 552;
Best Local Similarity 88.9%; Pred. No. 3.5e+02;
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 15 GAGCCCTTACAAAGACTATGAAGTAAA 41
|||||
362 GAGCCCTTACAAAGACTAAGACCAA 388

RESULT 74

```





POLYA=Yes.  
 FEATURES  
 source  
 Location/Qualifiers  
 1. .614  
 /organism="Rattus norvegicus"  
 /mol\_type="mRNA"  
 /strain="Sprague-Dawley"  
 /db\_xref="taxon:10116"  
 /clone="UI-R-CNI-cjg-g-03-0-UI"  
 /dev\_stage="adult"  
 /lab\_host="DH10B (Life Technologies)"  
 /clone\_lib="UI-R-CNI"  
 /note="Vector: p773D-PacI; Site\_1: Not I; Site\_2: Eco RI;  
 The UI-R-CNI library is a subtracted library derived from  
 the following pool of seven normalized rat libraries:  
 normalized rat seminal vesicles, normalized rat penis,  
 normalized rat bladder, normalized rat cervix, normalized  
 rat brown adipose, normalized rat fundus, and normalized  
 rat salivary gland. It was constructed according to the  
 procedure described by Bonaldo, Lennon & Soares (Genome  
 Research Genome 6: 791-806, 1996). For construction of  
 the CNI library, plasmid DNA from the pool of normalized  
 libraries was electroporated into competent bacteria for  
 the production of single-stranded circular DNA. This was  
 then used as a tracer in a subtractive hybridization with  
 a driver (PCR amplified inserts from a plasmid DNA template  
 preparation) comprising: a) a pool of about 34,000 clones  
 from the Rat Unigene Set corresponding to plates R-5-AA-NN  
 excluding plates R-5-MM and MN. This pool represented 40%  
 of the final driver population. b) a pool of about 29,000  
 clones from subtracted libraries CA0 and CAL corresponding  
 to plates R-CA0-AMV through R-CA0-AXS, R-CA0-AXZ through  
 R-CA0-BAZ, R-CA0-BFE through R-CA0-BHY, R-CA0-BJS,  
 R-CA0-BKE, R-CA0-BKG-H, R-CA0-BKJ-K, R-CA0-BKP through  
 R-CA0-BKS, R-CA0-BKU-V, R-CA0-BLY through R-CA0-BMA,  
 R-CA0-BMC through R-CA0-BME, R-CA0-BNS, R-CA0-BOB through  
 R-CA0-BOJ, R-CA0-BPA through R-CA0-BPG, R-CA0-BBA through  
 R-CA0-BDA, R-CA0-BHZ through R-CA0-BJF, R-CA0-BJR,  
 R-CA0-BJT through R-CA0-BKE, R-CA0-BKD, R-CA0-BKF,  
 R-CA0-BKI, R-CA0-BKT, R-CA0-BLF, R-CA0-BLH through  
 R-CA0-BLN, R-CA0-BLS, R-CA0-BLU-V, R-CA0-BNR, and  
 R-CA0-BLE. The resulting pool represented 20% of the  
 final driver population. c) a pool of about 15,000 clones  
 from non-normalized libraries CS0s, CT0s, CW0s, CX0s  
 and normalized libraries CS0, CT0, CU0, CW0, and CX0  
 corresponding to plates R-CS0s-CBD through R-CS0s-CBO,  
 R-CT0s-CCA through R-CT0s-CAX, R-CU0s-CBP through  
 R-CU0s-CCA, R-CW0s-CCB through R-CW0s-CGM, R-CX0s-CCN  
 through R-CX0s-CCX, R-CS0-BSD, R-CS0-BTD through  
 R-CS0-BTV, R-CS0-BVM, R-CT0-BTW through R-CT0-BUP,  
 R-CT0-BVN, R-CU0-BUQ through R-CU0-BVL, R-CW0-BVW through  
 R-CW0-BWP, R-CW0-BXN through R-CW0-BXO, R-CX0-BMQ through  
 R-CX0-BXM. The resulting pool represented 5% of the final  
 driver population. d) a pool of about 5,000 clones (1,000  
 from non-normalized eye library CV0 and 4,000 from  
 normalized eye library CV1) corresponding to plates  
 R-CV0-BRH through R-CV0-BRR, R-CV1-BRS through R-CV1-BSC,  
 R-CV1-BSE through R-CV1-BTC, and R-CV1-BVO through  
 R-CV1-BVU. This pool represented about 5% of the final  
 driver population. e) A pool of about 10,000 clones from  
 subtracted library BS2, BV0 and BV0p (7-9.5 kb cDNA  
 library fraction from rat whole embryo), and BX0 (0.5-7kb  
 cDNA library fraction from rat whole embryo) corresponding  
 to plates R-BS2-BDB through R-BS2-BFB, R-BV0-ANK through  
 R-BV0-ANR, R-BV0p-AOI through R-BV0p-AOX, and R-BX0-AOY  
 through R-BX0-ASH. The resulting pool represented 5% of  
 the final driver population. f) a pool of about 7,000  
 clones from the seven non-normalized libraries that make  
 up the tracer including CY0, CZ0, DA0, DB0, DC0, DD0, and  
 DE0 corresponding to plates R-CY0-BXP through R-CY0-BXZ,  
 R-CZ0-BYA through R-CZ0-BYI, R-CZ0-BZB-C, R-DA0-BYJ  
 through R-DA0-BYP, R-DA0-BZD through R-DA0-BZH, R-DB0-BYQ  
 through R-DB0-BZA, R-DC0-BZI through R-DC0-BZQ, R-DD0-BZC  
 through R-DD0-CBA, R-DD0-BZR through R-DD0-CAA,

R-DD0-CBB-C, and R-DE0-CAB through R-DE0-CAL. The  
 resulting pool represented about 10% of the final driver  
 population. g) a pool of about 2,000 clones from the pool  
 of normalized libraries, CNO, that makes up the tracer.  
 The corresponding plates are R-CNO-BKW through R-CNO-BLD,  
 R-CNO-BLG, R-CNO-BLP through R-CNO-BLR, R-CNO-BLT,  
 R-CNO-BLW-X, R-CNO-BMB, and R-CNO-BMF through R-CNO-BML.  
 This pool represented 5% of the final driver population.  
 h) a pool of the 28 most abundant clones in the CNO pool  
 corresponding to the following addresses: bkw-a-09-0-UI,  
 bkw-b-09-0-UI, bkw-b-11-0-UI, bkw-b-10-0-UI,  
 bkw-d-01-0-UI, bkw-d-06-0-UI, bkw-g-08-0-UI,  
 bkw-h-12-0-UI, bky-a-05-0-UI, bkr-a-06-0-UI,  
 bkr-a-11-0-UI, bkr-c-06-0-UI, bkr-c-09-0-UI,  
 bkr-d-10-0-UI, bla-a-01-0-UI, bla-a-02-0-UI,  
 bla-f-04-0-UI, bla-g-07-0-UI, bla-g-12-0-UI,  
 blb-a-12-0-UI, blb-f-02-0-UI, blc-a-11-0-UI,  
 blc-e-95-0-UI, bld-1-08-0-UI, bld-f-02-0-UI,  
 blg-h-04-0-UI, blr-a-03-0-UI, blt-f-08-0-UI. This pool  
 represented 5% of the final driver population. i) One  
 abundant CNO clone (corresponding to the address  
 bkr-a-11-0-UI) was digested with Not I and Eco RI and the  
 resulting insert was gel purified. This purified insert  
 was added directly to the driver so that it represented 5%  
 of the final driver population.  
 TAG TISSUE=cervix  
 TAG LIB=UI-R-CNI  
 TAG\_SEQ=GACCA"

ORIGIN  
 Query Match 52.9%; Score 22.2; DB 2; Length 614;  
 Best Local Similarity 88.9%; Pred.No. 3.5e+02;  
 Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 15 GAGCCCTTACAAAGACTATGAAGTAAA 41  
 |||||  
 Db 386 GAGCCCTTACAAAGACTATGAAGACCAA 412  
 |||||

RESULT 77  
 AA946356 620 bp mRNA linear EST 16-JUN-1998  
 LOCUS EST201855 Normalized rat lung, Bento Soares Rattus sp. cDNA clone  
 DEFINITION RLBH16 3' end, mRNA sequence.  
 AA946356  
 ACCESSION  
 VERSION AA946356.1 GI:3106272  
 KEYWORDS EST.  
 SOURCE Rattus sp.  
 ORGANISM Rattus sp.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridea; Muridae; Murinae; Rattus.  
 REFERENCE 1 (bases 1 to 620)  
 AUTHORS Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,  
 Kerlavage,A.R. and Adams,M.D.  
 TITLE Rat Genome Project: Generation of a Rat EST (EST) Catalog & Rat  
 Gene Index  
 JOURNAL Unpublished (1998)  
 COMMENT Contact: Lee, NH  
 The Institute for Genomic Research  
 9712, Medical Center Drive, Rockville, MD 20850, USA  
 Tel: (301)-838-3529  
 Fax: (301)-838-0208  
 Email: nhlee@tigr.org  
 Seq primer: M13-21.  
 Location/Qualifiers  
 1. .620  
 /organism="Rattus sp."  
 /mol\_type="mRNA"  
 /db\_xref="ATCC (inhost):2015290"  
 /db\_xref="taxon:10118"  
 /clone="RLBH16"  
 /clone\_lib="Normalized rat lung, Bento Soares"

/note="Organ: lung; Vector: pT7T3pac; Site\_1: EcoRI;  
Site\_2: NotI"

## ORIGIN

Query Match 52.9%; Score 22.2; DB 1; Length 620;  
Best Local Similarity 88.9%; Pred. No. 3.5e+02;  
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 15 GAGCCCTTACAAAGACTATGAAGTAAA 41  
|||||  
Db 338 GAGCCCTTACAAAGACTAAGAACCAAA 364  
|||||

## RESULT 78

CB326875 CB326875 673 bp mRNA linear EST 12-AUG-2004  
LOCUS UI-R-F50-crw-i-16-0-UI.s1 NCI\_CGAP\_F50 Rattus norvegicus cDNA clone  
DEFINITION IMAGE:7360314 3', mRNA sequence.

ACCESSION CB326875

VERSION CB326875

KEYWORDS EST

SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muroidea; Muridae; Murinae; Rattus.

1 (bases 1 to 673)

Bonaldo, M.F., Lennon, G. and Soares, M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

889548

Contact: Daniela S. Gerhard, Ph.D.

Office of Cancer Genomics

National Cancer Institute / NIH

Bldg. 31 Rm10A07 Bethesda, MD 20892

Email: cgapbs-remail.nih.gov

Oligo-dT track not found, Not I site shown in beginning of sequence

is likely internal to the message. cDNA Library Preparation: M.B.

Soares Lab Clone distribution: clones will be available through

IMAGE (http://image.llnl.gov)

Seq primer: M13 FORWARD

POLYA=Yes.

## FEATURES

source

1..673 Location/Qualifiers

/organism="Rattus norvegicus"

/mol\_type="mRNA"

/strain="Sprague-Dawley"

/db\_xref="taxon:10116"

/clones="IMAGE:7360314"

/dev\_stage="ADULT"

/lab\_host="DH10B (Life Technologies)"

/clone\_lib="NCI\_CGAP\_F50"

/note="Vector: pT7T3D-PacI; Site 1: Not I; Site 2: Eco RI;

UI-R-F50 is a cDNA library constructed from Swarm Rat

Chondrosarcoma. The library was constructed according to

Bonaldo, Lennon and Soares, Genome Research, 6:791-806,

1996. First strand cDNA synthesis was primed with an

oligo-dT primer containing a Not I site. Double stranded

cDNA was ligated to an EcoR I adaptor, digested with Not

I, and cloned directionally into pT7T3-Pac vector. The

oligonucleotide used to prime the synthesis of

first-strand cDNA contains a library tag sequence that is

located between the Not I site and the (dT)18 tail. The

sequence tag for this library is AGCCGCCGAT. The tumor was

provided by Dr Jeff Stevens from University of Iowa.

TAG\_TISSUE=Rat SRC-JWS Cell Line

TAG\_LIB=UI-R-F50

TAG\_SEQ=AGCCGCCGAT"

## ORIGIN

Query Match 52.9%; Score 22.2; DB 4; Length 673;  
Best Local Similarity 88.9%; Pred. No. 3.6e+02;

Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 15 GAGCCCTTACAAAGACTATGAAGTAAA 41  
|||||  
Db 374 GAGCCCTTACAAAGACTAAGAACCAAA 400  
|||||

## RESULT 79

CA504656 CA504656 689 bp mRNA linear EST 14-NOV-2002  
LOCUS UI-R-FJ0-cpw-h-08-0-UI.s1 UI-R-FJ0 Rattus norvegicus cDNA clone  
DEFINITION UI-R-FJ0-cpw-h-08-0-UI 3', mRNA sequence.

ACCESSION CA504656

VERSION CA504656.1

KEYWORDS GI:24995610

SOURCE EST.

ORGANISM Rattus norvegicus (Norway rat)

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muroidea; Muridae; Murinae; Rattus.

1 (bases 1 to 689)

Bonaldo, M.F., Lennon, G. and Soares, M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

889548

Contact: Soares, MB

Coordinated Laboratory for Computational Genomics

University of Iowa

375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: bento-soares@uiowa.edu

Tissue Procurement: Dr. James Lin, University of Iowa

cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa

cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Researchers may obtain clones from Research

Genetics (www.resgen.com).

The following repetitive elements were found in this cDNA

sequence: 1-30, >POLY A#Simple\_repeat (matched complement)

Seq primer: M13 FORWARD

POLYA=Yes.

## FEATURES

source

1..689 Location/Qualifiers

/organism="Rattus norvegicus"

/mol\_type="mRNA"

/strain="Sprague-Dawley"

/db\_xref="taxon:10116"

/clones="UI-R-FJ0-cpw-h-08-0-UI"

/tissue\_type="embryo"

/dev\_stage="embryo"

/lab\_host="DH10B (Life Technologies)"

/clone\_lib="UI-R-FJ0"

/note="Vector: pYX-Asc; Site 1: EcoR I; Site 2: Not I;

UI-R-FJ0 is a cDNA library containing the following

tissue(s): rat embryo. The library was constructed

according to Bonaldo, Lennon and Soares, Genome Research,

6:791-806, 1996. First strand cDNA synthesis was primed

with an oligo-dT primer containing a Not I site. Double

stranded cDNA was ligated to an EcoR I adaptor, digested

with Not I, and cloned directionally into pT7T3-Pac

vector. The oligonucleotide used to prime the synthesis of

first-strand cDNA contains a library tag sequence that is

located between the Not I site and the (dT)18 tail. The

sequence tag for this library is CATCTCTACT. This library

was created for the University of Iowa Program for Rat

Gene Discovery and Mapping (Val Sheffield, Bento Soares

and Tom Casavant)

TAG\_TISSUE=rat-embryo

TAG\_LIB=UI-R-FJ0

TAG\_SEQ=CATCTCTACT"

## ORIGIN

Query Match 52.9%; Score 22.2; DB 4; Length 689;  
 Best Local Similarity 88.9%; Pred. No. 3.6e+02;  
 Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 15 GAGCCCTTACAAAGACTATGAAGTAAA 41  
 |||||  
 Db 386 GAGCCCTTACAAAGACTAAGAACCAA 412

RESULT 80  
 LOCUS CK838771 736 bp mRNA linear EST 05-MAR-2004  
 DEFINITION UI-R-AC1-xx-c-10-0-UI.s10 UI-R-AC1 Rattus norvegicus cDNA clone  
 UI-R-AC1-xx-c-10-0-UI 3', mRNA sequence.

ACCESSION CK838771  
 VERSION  
 KEYWORDS

SOURCE Rattus norvegicus (Norway rat)  
 ORGANISM

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muroidae; Muridae; Murinae; Rattus.  
 1 (bases 1 to 736)

AUTHORS Bonaïdo, M.F., Lennon, G. and Soares, M.B.  
 TITLE Normalization and subtraction: two approaches to facilitate gene  
 discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)

PUBMED 8889548

COMMENT Contact: Soares, MB

Coordinated Laboratory for Computational Genomics  
 University of Iowa  
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: bento-soares@uiowa.edu

cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa

DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Distribution information can be found at

http://genome.uiowa.edu/distribution/rat.html

Seq primer: M13 FORWARD

POLYA=Yes.

FEATURES Location/Qualifiers

1..736  
 /organism="Rattus norvegicus"

/mol\_type="mRNA"

/strain="Sprague-Dawley"

/db\_xref="taxon:10116"

/clones="UI-R-AC1-xx-c-10-0-UI"

/dev\_stage="adult"

/lab\_host="DH10B (Life Technologies)"

/clone\_lib="UI-R-AC1"

/note="Vector: pT73D-PacI; Site 1: Not I; Site 2: Eco RI;

The UI-R-AC1 library is a normalized library constructed

from 16.5 dpc rat atrioventricular (AV) canal. The tag

is a string of 5 nucleotides present between the Not I

site and the oligo-dr track. The library was constructed

as described by Bonaïdo, Lennon and Soares, Genome

Research 6: 791-806, 1996. Tissue provided by Jim Lin,

Department of Biology, University of Iowa.

TAG\_TISSUE=muscle

TAG\_LIB=UI-R-AC1

TAG\_SEQ=AAG"

ORIGIN

Query Match 52.9%; Score 22.2; DB 5; Length 736;  
 Best Local Similarity 88.9%; Pred. No. 3.6e+02;  
 Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 15 GAGCCCTTACAAAGACTATGAAGTAAA 41  
 |||||  
 Db 374 GAGCCCTTACAAAGACTAAGAACCAA 400

RESULT 81  
 LOCUS C2983576

DEFINITION

197310 Tomato MboI BAC Library Lycopersicon esculentum genomic

clone SL\_MboI0086A08 5, genomic survey sequence.

ACCESSION C2983576

VERSION

KEYWORDS

SOURCE

ORGANISM

Lycopersicon esculentum (Solanum lycopersicum)

Eukaryota; Viridiplantae;

Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;

asterids; lamids; Solanales; Solanaceae; Lycopersicon.

1 (bases 1 to 859)

AUTHORS Mueller, L.A., Buel, R.M., Wang, Y., Tanksley, S.D., Giovannoni, J.J.,

Van Eck, J. and Stack, S.

TITLE BAC end sequencing from three Solanum lycopersicon libraries

COMMENT Unpublished (2005)

Other GSSs: 197309

Contact: Lukas Mueller

Tanksley Lab, Dept. of Plant Breeding

Cornell University

251 Emerson Hall, Ithaca, NY 14853, USA

Tel: 607-255-6557

Fax: 607-255-6683

Email: sgn-feedback@sgn.cornell.edu

Plate: 86 row: A column: 8

Seq primer: T7

Class: BAC ends

High quality sequence start: 59

High quality sequence stop: 316.

FEATURES

source

1..859

/organism="Lycopersicon esculentum"

/mol\_type="genomic DNA"

/cultivar="Heinz 1706"

/db\_xref="taxon:4081"

/clones="SL\_MboI0086A08"

/lab\_host="E. coli"

/clone\_lib="Tomato MboI BAC Library"

/note="Vector: pBelOBAC11; Site 1: MboI"

ORIGIN

Query Match 52.9%; Score 22.2; DB 13; Length 859;

Best Local Similarity 77.1%; Pred. No. 3.8e+02;

Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 5 AGCGTTGCCGCGCCCTTACAAAGACTATGAAGTA 39

|||||

Db 354 AGGCATGCCCAACCCCATCAAGATTCTGAAGTA 388

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||



```

/lab host="DH10B"
/clone_lib="RIKEN full-length enriched, 10 day neonate
skin"
/notes="Site 1: SalI; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5',
GAGAGAGAGAGATCCAGAGCTCTTTTTTTTTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 10.0 and subtraction to Rot = 100.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5', GAGAGAGATTCGAGTGAATTAATTAATCCCCCCCCCC
3']. cDNA was cloned into the XhoI and BamHI sites.
Vector: a modified pBluescript KS(+) after bulk excision
from Lambda FLC I"
```

## ORIGIN

```

Query Match          52.4%; Score 22; DB 1; Length 266;
Best Local Similarity 73.7%; Pred. No. 3.7e+02;
Matches 26; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 4 CACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAA 41
    ||||| ||||| ||||| ||||| ||||| |||||
Db 44 CCGGATTGGCAAGCTGTTCACAACTACTATTAAATAAA 81
```

## RESULT 84

```

CO755512
LOCUS      Mdfirt3049p06.y1 Mdfirt Malus x domestica cDNA clone Mdfirt3049p06 5'
DEFINITION similar to TR:Q9SQZ1 Q9SQZ1 PUTATIVE 40S RIBOSOMAL PROTEIN S17. ;,
mRNA sequence.
```

## ACCESSION

```

VERSION    CO755512.1 GI:50890761
```

## KEYWORDS

```

SOURCE     EST.
```

## ORGANISM

```

Malus x domestica
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids I; Rosales; Rosaceae; Maloideae; Malus.
```

## REFERENCE

## AUTHORS

```

1 (bases 1 to 381)
Korban, S., Vodkin, L., Liu, L., Gasic, K., Gonzales, O., Hernandez, A.,
Aldwinckle, H., Malnoy, M., Carroll, N., Goldsbrough, P., Orvis, K.,
Clifton, S., Pape, D., Marra, M., Hillier, L., Martin, J., Wylie, T.,
Dante, M., Theising, B., Bowers, Y., Gibbons, M., Ritter, E., Ronko, I.,
Teagareishvili, R., Kennedy, S., Waterston, R. and Wilson, R.
Apple Functional Genomics grant - NSF 0321702
Unpublished (2004)
```

## TITLE

## JOURNAL

## COMMENT

```

Contact: Schuyler S. Korban
Apple Functional Genomics grant - NSF 0321702
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
```

```

Email: est@watson.wustl.edu
```

```

Library materials provided by: Schuyler S. Korban Library
constructed by: K. Gasic Library sequenced by: Washington
University Genome Sequencing Center
```

```

WashU EST name: aa164h03.y1
```

```

Seq primer: -40UP from Gibco
```

```

High quality sequence stop: 284.
```

## FEATURES

## source

```

1..381
    /organism="Malus x domestica"
    /mol_type="mRNA"
    /culturvar="GoldRush"
    /db_xref="taxon:3750"
    /clone="Mdfirt3049p06"
```

```

/tissue_type="Fruit"
/lab_host="DH10B ampicillin resistant"
/clone_lib="Mdfirt"
/notes="Vector: pBluescript II SK (+); Site 1: NotI;
Site 2: EcoRI; Total RNA was extracted separately from
each stage [young fruitlet (<1cm), young fruitlet (1 cm
dia.), young fruitlet (12cm dia.), maturing fruit 1,
maturing fruit II, mature fruit], using the 'pine tree'
method. Poly(A)+mRNA was isolated twice from total RNA
from each stage using the Oligotex Direct mRNA kit
(Qiagen). mRNA was reverse transcribed into double
stranded cDNA using a modified oligo18(4T) primer with an
identifying tag sequence (see table below). cDNA's from
different stages were pooled in equal amounts before
adapter ligation. Tag identification when sequencing from
5' end: Stage 1 (young fruitlet) insert 18(A)TCGTG; Stage
2 (young fruitlet 1cm dia) insert 18(A)TCGTG; Stage 3
(young fruitlet 12cm dia) insert 18(A)TCGTG; Stage 4
(maturing fruit I) insert 18(A)TCGA; Stage 5 (maturing
fruit II) insert 18(A)TCGA; Stage 6 (mature fruit) insert
18(A)TCGTG; Tag identification when sequencing from 3'
end: Stage 1 (young fruitlet) CACGA18(T) insert; Stage 2
(young fruitlet 1cm dia) CAGCA18(T) insert; Stage 3 (young
fruitlet 12cm dia) ACCGA18(T) insert; Stage 4 (maturing
fruit I) TCGCA18(T) insert; Stage 5 (maturing fruit II)
TCGCA18(T) insert; Stage 6 (mature fruit) ACGCA18(T)
insert. Double stranded cDNAs were size selected (more
than 450 bp), adapted with EcoRI adapters at both ends
and then digested with NotI. The cDNAs were then
directionally cloned into EcoRI-NotI digested pBS II SK(+)
phagemid vector(Stratagene). Identification of adaptors
and tags in 5'-end sequenced clones:
(Vector) . . .TAGCCTT(End Vector)(Start
EcoRI adaptor)GATATCGAATTCATTTGTGTGGG (End
EcoRI adaptor)(Start Insert) . . .AAAAAAAAAAAAAAAAAAAA
Insert) (Start Tag)TCGA(End Tag) (Start
NotI site/Vector)GCGCGCCACCGCGG . . . The total number of
white colony forming units (cfu) in the primary library
before amplification was 2.1x10^6 cfu (colony forming
units). The background of empty clones was less than 1%.
Inserts ranged from 0.5kb to 4 kb, as determined by PCR.
Purified plasmid DNA from the primary library was
converted to single-stranded circles and used as a
template for PCR amplification using the T7 and T3 priming
sites flanking the cloned cDNA inserts. The purified PCR
products, representing the entire cloned cDNA population,
were used as a driver for normalization. Hybridization
between the single-stranded library and the PCR products
was carried out for 44 hours at 30C. Unhybridized
single-stranded DNA circles were separated from hybridized
DNA rendered partially double-stranded and electroporated
into DH10B cells to generate the normalized library. The
total number of clones with insert was 5.6x10^6 cfu.
Background of empty clones was less than 1%."
```

## ORIGIN

```

Query Match          52.4%; Score 22; DB 8; Length 381;
Best Local Similarity 73.7%; Pred. No. 3.9e+02;
Matches 28; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 3 CCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAA 40
    ||||| ||||| ||||| ||||| ||||| |||||
Db 59 CCACATGGTTCGCTCGTACAAAGACAGTGAAGAAA 96
```

## RESULT 85

## LOCUS

## DEFINITION

```

Mdfirt3038j23.y1 Mdfirt Malus x domestica cDNA clone Mdfirt3038j23 5'
```

```

similar to TR:Q9SWF6 Q9SWF6 40S RIBOSOMAL PROTEIN S17. ;, mRNA
```

```

sequence.
```

```

CO541025
```

```

ACCESSION
```

```

VERSION    CO541025.1 GI:50353074
```

```

CO541025          434 bp      mRNA      linear      EST 16-JUL-2004
Mdfirt3038j23.y1 Mdfirt Malus x domestica cDNA clone Mdfirt3038j23 5'
similar to TR:Q9SWF6 Q9SWF6 40S RIBOSOMAL PROTEIN S17. ;, mRNA
sequence.
```







constructed by: K. Gasic Library sequenced by: Washington  
University Genome Sequencing Center  
WashU EST name: aa06b02.y1  
Seq primer: -40UP from Gibco  
High quality sequence stop: 461.  
Location/Qualifiers

## FEATURES

source

```

1..482
/organism="Malus x domestica"
/mol_type="mRNA"
/cultivar="GoldRush"
/db_xref="taxon:3750"
/clone="Mdb616C03"
/tissue_type="Shoot internodes"
/lab_host="DH10B ampicillin resistant"
/clone_lib="Mdbst"
/note="Vector: pBluescript II SK (+); Site 1: NotI;
Site 2: EcoRI; Total RNA was extracted separately from
each stage [dormant shoot internodes, active shoot
internodes, actively-growing shoot internodes], using the
'pine tree' method. Poly(A)-mRNA was isolated twice from
total RNA from each stage using the Oligotex Direct mRNA
kit (Qiagen). mRNA was reverse transcribed into double
stranded cDNA using a modified oligo18(dT) primer with an
identifying tag sequence (see table below). cDNA's from
different stages were pooled in equal amounts before
adaptor ligation. Tag identification when sequencing from
5' end: Stage 1 (dormant shoot internodes) insert
18(A)TCGTG; Stage 2 (active shoot internodes) insert
18(A)TCGTG; Stage 3 (actively-growing shoot internodes)
insert 18(A)TCGTG; Tag identification when sequencing from
3' end: Stage 1 (dormant shoot internodes) CACGA18(T)
insert; Stage 2 (active shoot internodes) CACGA18(T)
insert; Stage 3 (actively-growing shoot internodes)
ACCGA18(T) insert; Double stranded cDNAs were size
selected (more than 450 bp), adapted with EcoRI adaptors
at both ends and then digested with NotI. The cDNAs were
then directionally cloned into EcoRI-NotI digested pBS II
SK(+) phagemid vector(Stratagene). Identification of
adaptors and tags in 5'-end sequenced clones:
(Vector)...TAAGCTT(End Vector)(Start
EcoRI adaptor)GATATCGAATTCATTGTGTGGG (End
Insert) (Start Tag)TCGA(End Tag)(Start
NotI site/Vector)GCGCGCCACCGGG... The total number of
white colony forming units (cfu) in the primary library
before amplification was 2.7x107 cfu (colony forming
units). The background of empty clones was less than 2%.
Purified plasmid DNA from the primary library was
converted to single-stranded circles and used as a
template for PCR amplification using the T7 and T3 priming
sites flanking the cloned cDNA inserts. The purified PCR
products, representing the entire cloned cDNA population,
were used as a driver for normalization. Hybridization
between the single-stranded library and the PCR products
was carried out for 44 hours at 30C. Unhybridized
single-stranded DNA circles were separated from hybridized
DNA rendered partially double-stranded and electroporated
into DH10B cells to generate the normalized library. The
total number of clones with insert was 8x105 cfu.
Background of empty clones was less than 1%."

```

## ORIGIN

Query Match 52.4%; Score 22; DB 8; Length 482;  
Best Local Similarity 73.7%; Pred. NO. 4.le+02;  
Matches 28; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Oy 3 CCACGCTTGGCGGACCCCTTACAAGACTATGAGCTAA 40  
|||||  
Db 61 CCACATCGGTGCGGTCCGTCAAAAGACAGTGAGAAA 98  
|||||

RESULT 89

## CV084234

LOCUS  
DEFINITION

CV084234 487 bp mRNA linear EST 26-AUG-2004  
Mdb5027m14.y1 Mdb Malus x domestica cDNA clone Mdb5027m14.5,  
similar to TR:Q9SWF6 Q9SWF6 40S RIBOSOMAL PROTEIN S17.1; mRNA  
sequence.

## ACCESSION

VERSION

CV084234

EST.

CV084234.1

GI:51563489

Malus x domestica

Malus x domestica

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 487)  
Korban, S., Vodka, L., Liu, L., Gasic, K., Gonzales, O., Hernandez, A.,  
Alwinckle, H., Malnoy, M., Carroll, N., Goldsbrough, P., Orvis, K.,  
Clifton, S., Pape, D., Marra, M., Hillier, L., Martin, J., Wylie, T.,  
Dante, M., Theising, B., Bowers, Y., Gibbons, M., Ritter, B., Ronko, I.,  
Tsagarishvili, R., Kennedy, S., Waterston, R. and Wilson, R.  
Apple Functional Genomics grant - NSF 0321702  
Unpublished (2004)

## TITLE

JOURNAL

COMMENT

Contact: Schuyler S. Korban  
Apple Functional Genomics grant - NSF 0321702  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu

Library materials provided by: Schuyler S. Korban Library  
constructed by: K. Gasic Library sequenced by: Washington  
University Genome Sequencing Center  
WashU EST name: aa50g07.y1

Seq primer: -40UP from Gibco

High quality sequence stop: 487.

## FEATURES

source

```

1..487
/organism="Malus x domestica"
/mol_type="mRNA"
/cultivar="GoldRush"
/db_xref="taxon:3750"
/clone="Mdb5027m14"
/tissue_type="Bud"
/lab_host="DH10B ampicillin resistant"
/clone_lib="Mdbb"
/note="Vector: pBluescript II SK (+); Site 1: NotI;
Site 2: EcoRI; Total RNA was extracted separately from
each stage [dormant terminal bud, dormant lateral bud,
active lateral bud], using the 'pine tree' method.
Poly(A)-mRNA was isolated twice from total RNA from each
stage using the Oligotex Direct mRNA kit (Qiagen). mRNA
was reverse transcribed into double stranded cDNA using a
modified oligo18(dT) primer with an identifying tag
sequence (see table below). cDNAs from different stages
were pooled in equal amounts before adaptor ligation. Tag
identification when sequencing from 5' end: Stage 1
(dormant terminal bud) insert 18(A)TCGTG; Stage 2 (dormant
lateral bud) insert 18(A)TCGTG; Stage 3 (active lateral
bud) insert 18(A)TCGTG; Tag identification when sequencing
from 3' end: Stage 1 (dormant terminal bud) CACGA18(T)
insert; Stage 2 (dormant lateral bud) CACGA18(T) insert;
Stage 3 (active lateral bud) ACCGA18(T) insert; Double
stranded cDNAs were size selected (more than 450 bp),
adapted with EcoRI adaptors at both ends and then
digested with NotI. The cDNAs were then directionally
cloned into EcoRI-NotI digested pBS II SK(+) phagemid
vector(Stratagene). Identification of adaptors and tags in
5'-end sequenced clones: (Vector)...TAAGCTT(End
Vector)(Start EcoRI adaptor)GATATCGAATTCATTGTGTGGG (End
EcoRI adaptor)(Start Insert).....AAAAAAAAAAAAAAAAAAAAA(End
Insert) (Start Tag)TCGA(End Tag)(Start
NotI site/Vector)GCGCGCCACCGGG... The total number of
white colony forming units (cfu) in the primary library
before amplification was 4x105 cfu (colony forming

```







Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homiidae; Homo.  
 1 (bases 1 to 703)  
 NIH-MGC <http://mgs.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Daniela S. Gerhard, Ph.D.  
 Office of Cancer Genomics  
 National Cancer Institute / NIH  
 Bldg. 31 Em10A07 Bethesda, MD 20892  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: Mary Hendrix  
 cDNA Library Preparation: M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: NDAM1164 row: 1 column: 08  
 High quality sequence start: 273  
 High quality sequence stop: 378.  
 Location/Qualifiers  
 source  
 1..703  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:30923599"  
 /tissue\_type="Chondrosarcoma Lung Metastasis cell lines"  
 /lab\_host="DH10B (T1 phage resistant)"  
 /clone\_lib="NIH\_MGC\_212"  
 /note="Organ: Lung; Vector: pYX-Asc; Site 1: EcoR I;  
 Site 2: Not I; The library was constructed according  
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
 1996. Denatured RNA was size fractionated on a 1% agarose  
 gel. First strand cDNA synthesis was primed with oligo-dT  
 primer containing a Not I site. Double strand cDNA was  
 size selected according to mRNA size fraction, ligated  
 with EcoR I adaptor, digested with Not I and then cloned  
 directionally into pYX-Asc vector. The library tag  
 sequence located between the Not I site and the polyA tail  
 is GATAGGCCA. Tissue was provided by Mary Hendrix."

# FEATURES source

Query Match 52.4%; Score 22; DB 8; Length 703;  
 Best Local Similarity 73.7%; Pred. No. 4.4e+02;  
 Matches 28; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
 Oy 1 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGT 38  
 |||||  
 Db 328 GCCCTTGCCTGCTCGCGCCCTACACAGACTATGTGTT 365  
 |||||  
 RESULT 97  
 B1561919  
 LOCUS  
 DEFINITION  
 603254822F1 NIH\_MGC\_97 Homo sapiens cDNA clone IMAGE:5297404 5',  
 mRNA sequence.  
 B1561919  
 VERSION  
 B1561919.1 GI:15449233  
 EST.  
 SOURCE  
 Homo sapiens (human)  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homiidae; Homo.  
 1 (bases 1 to 745)  
 NIH-MGC <http://mgs.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki

## ORIGIN

Query Match 52.4%; Score 22; DB 8; Length 703;  
 Best Local Similarity 73.7%; Pred. No. 4.4e+02;  
 Matches 28; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
 Oy 1 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGT 38  
 |||||  
 Db 328 GCCCTTGCCTGCTCGCGCCCTACACAGACTATGTGTT 365  
 |||||

RESULT 97  
 B1561919  
 LOCUS  
 DEFINITION  
 603254822F1 NIH\_MGC\_97 Homo sapiens cDNA clone IMAGE:5297404 5',  
 mRNA sequence.  
 B1561919  
 VERSION  
 B1561919.1 GI:15449233  
 EST.  
 SOURCE  
 Homo sapiens (human)  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homiidae; Homo.  
 1 (bases 1 to 745)  
 NIH-MGC <http://mgs.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki

Toshiyuki and Piero Carninci (RIKEN)  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLAM11752 row: k column: 05  
 High quality sequence stop: 667.  
 Location/Qualifiers  
 source  
 1..745  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5297404"  
 /lab\_host="DH10B"  
 /clone\_lib="NIH\_MGC\_97"  
 /note="Organ: testis; Vector: pBluescriptR (modified  
 pBluescript KS+); Site\_1: BamHI; Site\_2: SalI-XhoI  
 (gtcgag); Oligo-dT primed using primer  
 5'-TTTTTTTTTTTTTTVN-3', size-selected for average  
 insert size 2.2 kb and normalized to 200. This is a  
 primary library enriched for full-length clones and  
 constructed using the Cap-trapper method (Carninci, in  
 preparation). Library constructed by M. Brownstein  
 (NHGRI/NHGRI, National Institutes of Health). Note: this is  
 a NIH\_MGC Library."

# FEATURES source

Query Match 52.4%; Score 22; DB 2; Length 745;  
 Best Local Similarity 73.7%; Pred. No. 4.4e+02;  
 Matches 28; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
 Oy 5 ACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
 |||||  
 Db 608 AAGTTTCACAGAGCTTCAAAGACTCTGAAGAAAG 645  
 |||||

## ORIGIN

Query Match 52.4%; Score 22; DB 2; Length 745;  
 Best Local Similarity 73.7%; Pred. No. 4.4e+02;  
 Matches 28; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
 Oy 5 ACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
 |||||  
 Db 608 AAGTTTCACAGAGCTTCAAAGACTCTGAAGAAAG 645  
 |||||

# RESULT 98 B2436965

LOCUS  
 DEFINITION  
 B2436965 BONEV45TF BO\_1.6\_2\_KB\_tot Brassica oleracea genomic clone BONEV45,  
 genomic survey sequence.  
 B2436965  
 VERSION  
 B2436965.1 GI:26689633  
 GSS.  
 KEYWORDS  
 SOURCE  
 Brassica oleracea  
 ORGANISM  
 Brassica oleracea  
 Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
 rosids; eurosids II; Brassicales; Brassicaceae; Brassica.  
 1 (bases 1 to 747)  
 Ayele,M., Haas,B.J., Kumar,N., Wu,H., Xiao,Y., Van Aken,S.,  
 Utterback,T.R., Wortman,J.R., White,O.R. and Town,C.D.  
 Whole genome shotgun sequencing of Brassica oleracea and its  
 application to gene discovery and annotation in Arabidopsis  
 Genome Res. 15 (4), 487-495 (2005)  
 15805490  
 Other\_GSSs: BONEV45TR  
 Contact: Chris Town  
 TIGR  
 9712 Medical Center Drive, Rockville, MD 20850, USA.  
 Tel: 301-838-3523  
 Fax: 301-838-0208  
 Email: cdtown@tigr.org  
 DNA is from a doubled haploid provided by Tom Osborn.  
 Seq primer: TF  
 Class: sheared ends.  
 Location/Qualifiers  
 source  
 1..747  
 /organism="Brassica oleracea"  
 /mol\_type="genomic DNA"  
 /strain="T01000DH3"  
 /db\_xref="taxon:3712"

# FEATURES source

Query Match 52.4%; Score 22; DB 2; Length 745;  
 Best Local Similarity 73.7%; Pred. No. 4.4e+02;  
 Matches 28; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
 Oy 5 ACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
 |||||  
 Db 608 AAGTTTCACAGAGCTTCAAAGACTCTGAAGAAAG 645  
 |||||

1 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGT 38  
664 GCTCTTGCTCGTCTGCGCCCATACACAGACTATGTAGT 701

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2006, 15:21:38 ; Search time 94.15 Seconds  
(without alignments)

834.695 Million cell updates/sec

Title: US-10-600-816-30

Perfect score: 42

Sequence: 1 gccacgctggccgagcc.....acaaagactgaagtaag 42

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1403666 seqs, 935554401 residues

Total number of hits satisfying chosen parameters: 2807332

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 100% summaries

Database : Issued Patents NA:\*

- 1: /EMC Celerra\_SID33/ptodata/2/ina/1 COMB.seq:\*
- 2: /EMC Celerra\_SID33/ptodata/2/ina/5 COMB.seq:\*
- 3: /EMC Celerra\_SID33/ptodata/2/ina/6A COMB.seq:\*
- 4: /EMC Celerra\_SID33/ptodata/2/ina/6B COMB.seq:\*
- 5: /EMC Celerra\_SID33/ptodata/2/ina/7 COMB.seq:\*
- 6: /EMC Celerra\_SID33/ptodata/2/ina/H COMB.seq:\*
- 7: /EMC Celerra\_SID33/ptodata/2/ina/PTUS COMB.seq:\*
- 8: /EMC Celerra\_SID33/ptodata/2/ina/PP COMB.seq:\*
- 9: /EMC Celerra\_SID33/ptodata/2/ina/RE COMB.seq:\*
- 10: /EMC Celerra\_SID33/ptodata/2/ina/backfiles.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Query Length	ID	Description
1	42	100.0	601	3	US-09-949-016-29173 Sequence 29173, A
2	42	100.0	601	3	US-09-949-016-126957 Sequence 126957, A
3	42	100.0	1212	3	US-09-188-930-249 Sequence 249, App
4	42	100.0	1212	3	US-09-312-283C-249 Sequence 249, App
5	42	100.0	1228	3	US-09-495-050A-223 Sequence 223, App
6	42	100.0	2289	3	US-09-949-016-2613 Sequence 2613, App
7	42	100.0	2456	3	US-09-949-016-572 Sequence 572, App
8	42	100.0	26086	3	US-09-949-016-15355 Sequence 15355, A
9	42	100.0	26238	3	US-09-949-016-12314 Sequence 12314, A
10	22.8	54.3	1500	3	US-09-555-704-3 Sequence 3, Appli
11	22.8	54.3	4512	2	US-08-224-391-52 Sequence 52, Appli
12	22.8	54.3	4512	2	US-08-484-304-52 Sequence 52, Appli
13	21.6	51.9	19319	3	US-09-949-016-14107 Sequence 14107, A
c 14	21.6	51.4	452	3	US-09-513-999C-20939 Sequence 20939, A
15	21.6	51.4	601	3	US-09-949-016-177622 Sequence 177622, A
16	21.6	51.4	601	3	US-09-949-016-177623 Sequence 177623, A
c 17	21.6	51.4	70000	3	US-09-851-896-3 Sequence 3, Appli
c 18	21.6	51.4	76399	3	US-09-949-016-16819 Sequence 16819, A
19	21.4	51.0	407	3	US-09-513-999C-32761 Sequence 32761, A
c 20	21	50.0	136264	3	US-09-949-016-12756 Sequence 12756, A
c 21	21	50.0	136265	3	US-09-949-016-13001 Sequence 13001, A
c 22	20.8	49.5	1728	3	US-09-107-532A-3086 Sequence 3086, App
c 23	20.8	49.5	100877	3	US-09-949-016-13276 Sequence 13276, A

c	24	20.8	49.5	140725	3	US-09-949-016-17074 Sequence 17074, A
c	25	20.6	49.0	2772	3	US-09-949-016-1950 Sequence 1950, App
c	26	20.6	49.0	36895	3	US-09-949-016-13692 Sequence 13692, A
c	27	20.4	48.6	264	3	US-09-702-705-823 Sequence 823, App
c	28	20.4	48.6	264	3	US-09-736-457-823 Sequence 823, App
c	29	20.4	48.6	264	3	US-09-614-1248-823 Sequence 823, App
c	30	20.4	48.6	264	3	US-09-671-325-823 Sequence 823, App
c	31	20.4	48.6	264	3	US-09-589-184-823 Sequence 823, App
c	32	20.4	48.6	264	3	US-09-558-824-823 Sequence 823, App
c	33	20.4	48.6	264	3	US-10-017-754-823 Sequence 823, App
c	34	20.4	48.6	264	3	US-09-651-563-823 Sequence 823, App
c	35	20.4	48.6	363	3	US-09-902-540-1194 Sequence 5294, App
c	36	20.4	48.6	707	5	US-09-974-300-5294 Sequence 5294, App
c	37	20.4	48.6	708	5	US-09-974-300-5134 Sequence 5134, App
c	38	20.4	48.6	2523	4	US-10-094-749-1277 Sequence 1277, App
c	39	20.4	48.6	2648	3	US-10-104-047-349 Sequence 349, App
c	40	20.4	48.6	21295	3	US-09-902-540-1194 Sequence 1194, App
c	41	20	47.6	576	2	US-08-086-428B-49 Sequence 49, Appl
c	42	20	47.6	576	2	US-08-468-570-49 Sequence 49, Appl
c	43	20	47.6	576	2	US-08-290-665A-49 Sequence 49, Appl
c	44	20	47.6	576	3	US-08-466-601A-49 Sequence 49, Appl
c	46	20	47.6	211049	7	PCT-US95-10398-49 Sequence 49, Appl
c	47	19.8	47.1	546	3	US-09-949-016-15770 Sequence 15770, A
c	48	19.8	47.1	601	3	US-09-949-016-148482 Sequence 148482, A
c	49	19.8	47.1	601	3	US-09-949-016-148483 Sequence 148483, A
c	50	19.8	47.1	1545	3	US-09-912-484-1 Sequence 1, Appli
c	51	19.8	47.1	1545	3	US-09-912-484-1 Sequence 1, Appli
c	52	19.8	47.1	8961	3	US-09-949-016-15910 Sequence 15910, A
c	53	19.8	47.1	9382	3	US-09-949-016-15306 Sequence 15306, A
c	54	19.8	47.1	75212	3	US-09-949-016-13313 Sequence 13313, A
c	55	19.8	47.1	75212	3	US-09-949-016-13314 Sequence 13314, A
c	56	19.8	47.1	75212	3	US-09-949-016-13315 Sequence 13315, A
c	57	19.8	47.1	87523	3	US-09-949-016-12670 Sequence 12670, A
c	58	19.8	47.1	87523	3	US-09-949-016-15047 Sequence 15047, A
c	59	19.8	47.1	87523	3	US-09-949-016-15048 Sequence 15048, A
c	60	19.8	47.1	87523	3	US-09-949-016-15049 Sequence 15049, A
c	61	19.8	47.1	87869	3	US-09-949-016-11744 Sequence 11744, A
c	62	19.8	47.1	87869	3	US-09-949-016-15044 Sequence 15044, A
c	63	19.8	47.1	87869	3	US-09-949-016-15045 Sequence 15045, A
c	64	19.8	47.1	87869	3	US-09-949-016-15046 Sequence 15046, A
c	65	19.8	47.1	199471	3	US-09-949-016-14083 Sequence 14083, A
c	66	19.6	46.7	337	3	US-08-956-171B-4323 Sequence 4323, App
c	67	19.6	46.7	337	3	US-08-781-986A-4323 Sequence 4323, App
c	68	19.6	46.7	601	3	US-09-949-016-146770 Sequence 146770, A
c	69	19.6	46.7	1020	3	US-09-205-815B-49 Patent No. 5486473
c	70	19.6	46.7	1500	10	US-08-956-171B-316 Sequence 316, App
c	71	19.6	46.7	9310	3	US-08-781-986A-316 Sequence 316, App
c	72	19.6	46.7	9310	3	US-09-452-638-52 Sequence 52, Appl
c	73	19.6	46.7	10892	3	US-09-452-638-52 Sequence 52, Appl
c	74	19.6	46.7	10892	3	US-09-121-587A-12 Sequence 12, Appl
c	75	19.6	46.7	36677	3	US-09-949-002-832 Sequence 832, App
c	76	19.6	46.7	96340	3	US-09-949-016-15863 Sequence 15863, A
c	77	19.6	46.7	132456	3	US-09-949-016-13750 Sequence 13750, A
c	78	19.6	46.7	199495	3	US-09-949-016-13750 Sequence 13750, A
c	79	19.6	46.7	199495	3	US-09-949-016-15436 Sequence 15436, A
c	80	19.6	46.7	302604	3	US-09-949-016-14588 Sequence 14588, A
c	81	19.6	46.7	302604	3	US-09-949-016-14589 Sequence 14589, A
c	82	19.6	46.7	308362	3	US-09-949-016-17119 Sequence 17119, A
c	83	19.4	46.2	461	3	US-09-621-976-16747 Sequence 16747, A
c	84	19.4	46.2	601	3	US-09-949-016-95958 Sequence 95958, A
c	85	19.4	46.2	601	3	US-09-949-016-188486 Sequence 188486, A
c	86	19.4	46.2	938	3	US-09-270-767-12087 Sequence 12087, A
c	87	19.4	46.2	158735	3	US-09-949-016-11989 Sequence 11989, A
c	88	19.4	46.2	158735	3	US-09-949-016-17130 Sequence 17130, A
c	89	19.4	46.2	387902	3	US-09-949-016-14543 Sequence 14543, A
c	90	19.4	46.2	421883	3	US-09-949-016-12557 Sequence 12557, A
c	91	19.2	45.7	215	3	US-09-621-976-16083 Sequence 16083, A
c	92	19.2	45.7	450	3	US-09-252-991A-8407 Sequence 8407, App
c	93	19.2	45.7	579	3	US-09-878-281A-135 Sequence 135, App
c	94	19.2	45.7	579	3	US-09-878-281A-135 Sequence 135, App
c	95	19.2	45.7	580	3	US-09-878-281A-45 Sequence 45, Appl
c	96	19.2	45.7	580	3	US-09-878-281A-47 Sequence 47, Appl





c 243	18.8	44.8	222691	3	US-09-949-016-11762	Sequence 11762, A	316	18.2	43.3	1275	3	US-09-328-925-80	Sequence 80, Appl
c 244	18.8	44.8	222697	3	US-09-949-016-15842	Sequence 15842, A	317	18.2	43.3	1276	3	US-09-328-925-81	Sequence 81, Appl
c 245	18.8	44.8	235064	3	US-09-949-016-15390	Sequence 13390, A	318	18.2	43.3	1427	3	US-09-577-657A-2	Sequence 2, Appl
c 246	18.8	44.8	312470	3	US-09-949-016-14043	Sequence 14043, A	319	18.2	43.3	1427	3	US-09-577-657A-3	Sequence 3, Appl
c 247	18.8	44.8	336024	3	US-09-949-016-12373	Sequence 12373, A	c 320	18.2	43.3	1472	3	US-10-101-464A-966	Sequence 966, App
c 248	18.8	44.8	421118	3	US-09-949-016-16297	Sequence 16297, A	321	18.2	43.3	2340	3	US-08-742-877-3	Sequence 3, Appl
c 249	18.6	44.3	354	3	US-09-513-999C-13634	Sequence 13634, A	322	18.2	43.3	2775	3	US-09-053-871A-22	Sequence 22, Appl
c 250	18.6	44.3	579	3	US-09-878-281A-179	Sequence 179, App	323	18.2	43.3	2792	3	US-09-118-748-1	Sequence 1, Appl
c 251	18.6	44.3	601	3	US-09-949-016-32925	Sequence 32925, A	324	18.2	43.3	2802	3	US-08-742-877-1	Sequence 1, Appl
c 252	18.6	44.3	601	3	US-09-949-016-32926	Sequence 32926, A	325	18.2	43.3	2804	3	US-08-742-877-1	Sequence 2, Appl
c 253	18.6	44.3	601	3	US-09-949-016-32927	Sequence 32927, A	326	18.2	43.3	2833	3	US-09-949-016-5252	Sequence 5252, Ap
c 254	18.6	44.3	601	3	US-09-949-016-185573	Sequence 185573, A	327	18.2	43.3	7260	3	US-09-917-254-31	Sequence 31, Appl
c 255	18.6	44.3	601	3	US-09-949-016-185574	Sequence 185574, A	328	18.2	43.3	7260	3	US-09-919-497-24	Sequence 24, Appl
c 256	18.6	44.3	601	3	US-09-949-016-185575	Sequence 185575, A	329	18.2	43.3	7260	4	US-09-880-107-3739	Sequence 3739, Ap
c 257	18.6	44.3	601	3	US-09-949-016-201189	Sequence 201189, A	330	18.2	43.3	12570	3	US-09-949-002-850	Sequence 850, App
c 258	18.6	44.3	603	5	US-09-974-300-798	Sequence 798, App	331	18.2	43.3	20407	3	US-09-949-002-866	Sequence 66, App
c 259	18.6	44.3	1500	3	US-09-555-704-1	Sequence 1, Appl	332	18.2	43.3	24333	5	US-09-984-429-351	Sequence 351, App
c 260	18.6	44.3	54986	3	US-09-949-016-16716	Sequence 16716, A	333	18.2	43.3	24333	5	US-09-984-429-351	Sequence 460, App
c 261	18.6	44.3	68035	3	US-09-949-016-16219	Sequence 16219, A	c 334	18.2	43.3	27465	3	US-09-949-016-16561	Sequence 16561, A
c 262	18.6	44.3	87594	3	US-09-949-016-12135	Sequence 12135, A	335	18.2	43.3	36755	3	US-09-949-016-16594	Sequence 16594, A
c 263	18.6	44.3	87611	3	US-09-949-016-16139	Sequence 16139, A	336	18.2	43.3	38059	3	US-09-328-925-4	Sequence 4, Appl
c 264	18.6	44.3	96583	4	US-09-531-120-203	Sequence 203, App	337	18.2	43.3	38059	4	US-09-880-107-2125	Sequence 2125, Ap
c 265	18.6	44.3	103987	3	US-09-949-016-12513	Sequence 12513, A	c 338	18.2	43.3	67002	3	US-09-949-016-16803	Sequence 16803, A
c 266	18.6	44.3	103988	3	US-09-949-016-17050	Sequence 17050, A	c 339	18.2	43.3	71815	3	US-09-949-016-12501	Sequence 12501, A
c 267	18.6	44.3	111677	3	US-09-949-016-16946	Sequence 16946, A	c 340	18.2	43.3	119930	3	US-09-949-016-12677	Sequence 12677, A
c 268	18.6	44.3	198632	3	US-09-949-016-12781	Sequence 12781, A	c 341	18.2	43.3	119931	3	US-09-949-016-16319	Sequence 16319, A
c 269	18.6	44.3	198632	3	US-09-949-016-17393	Sequence 17393, A	c 342	18.2	43.3	125356	3	US-09-949-016-14186	Sequence 14186, A
c 270	18.4	43.8	300	4	US-09-237-648-2393	Sequence 2393, Ap	c 343	18.2	43.3	136058	3	US-09-949-016-12565	Sequence 12565, A
c 271	18.4	43.8	576	2	US-08-086-428B-48	Sequence 48, Appl	c 344	18.2	43.3	136480	3	US-09-949-016-17064	Sequence 17064, A
c 272	18.4	43.8	576	2	US-08-468-570-48	Sequence 48, Appl	c 345	18.2	43.3	141560	3	US-09-949-016-16476	Sequence 16476, A
c 273	18.4	43.8	576	2	US-08-290-665A-48	Sequence 48, Appl	c 346	18.2	43.3	161652	3	US-09-497-855A-40	Sequence 40, Appl
c 274	18.4	43.8	576	3	US-08-466-601A-48	Sequence 48, Appl	c 347	18.2	43.3	611587	4	US-09-531-120-209	Sequence 209, App
c 275	18.4	43.8	576	7	PCT-US95-10398-A	Sequence 48, Appl	c 348	18.2	43.3	4403765	3	US-09-103-840A-2	Sequence 2, Appl
c 276	18.4	43.8	601	3	US-09-949-016-54878	Sequence 54878, A	c 349	18.2	43.3	4411529	3	US-09-103-840A-2	Sequence 1, Appl
c 277	18.4	43.8	721	3	US-09-533-559-2071	Sequence 2071, Ap	c 350	18.2	42.9	142	3	US-08-956-171E-4735	Sequence 4735, Ap
c 278	18.4	43.8	830	4	US-09-237-648-3596	Sequence 3596, Ap	c 351	18.2	42.9	142	3	US-08-781-986A-4735	Sequence 4735, Ap
c 279	18.4	43.8	1629	3	US-09-967-552A-27	Sequence 27, Appl	c 352	18.2	42.9	400	3	US-08-956-171E-3969	Sequence 3969, Ap
c 280	18.4	43.8	1749	3	US-09-967-552A-3	Sequence 5, Appl	c 353	18.2	42.9	400	3	US-08-781-986A-3969	Sequence 3969, Ap
c 281	18.4	43.8	1794	3	US-09-967-552A-5	Sequence 3, Appl	c 354	18.2	42.9	601	3	US-09-949-016-19854	Sequence 19854, A
c 282	18.4	43.8	2025	3	US-09-540-236-1626	Sequence 1626, Ap	c 355	18.2	42.9	601	3	US-09-949-016-118117	Sequence 118117, A
c 283	18.4	43.8	2446	3	US-09-731-166-9	Sequence 9, Appl	c 356	18.2	42.9	636	3	US-08-956-171E-3664	Sequence 3664, Ap
c 284	18.4	43.8	6336	4	US-09-880-107-1537	Sequence 9, Appl	c 357	18.2	42.9	636	3	US-08-781-986A-3664	Sequence 3664, Ap
c 285	18.4	43.8	6373	3	US-08-956-171E-227	Sequence 227, App	c 358	18.2	42.9	685	4	US-09-605-703B-1873	Sequence 1873, Ap
c 286	18.4	43.8	6373	3	US-08-781-986A-227	Sequence 227, App	c 359	18.2	42.9	949	3	US-09-270-767-15170	Sequence 15170, A
c 287	18.4	43.8	7525	3	US-09-967-552A-72	Sequence 72, Appl	c 360	18.2	42.9	1752	3	US-09-270-767-12078	Sequence 12078, A
c 288	18.4	43.8	24533	3	US-09-949-016-16901	Sequence 16901, A	c 361	18.2	42.9	1752	3	US-09-248-796A-6141	Sequence 6141, Ap
c 289	18.4	43.8	30271	3	US-09-949-016-12796	Sequence 12796, A	c 362	18.2	42.9	2725	3	US-10-054-678-1	Sequence 1, Appl
c 290	18.4	43.8	30272	3	US-09-949-016-14006	Sequence 14006, A	c 363	18.2	42.9	2833	3	US-09-270-767-15279	Sequence 15279, A
c 291	18.4	43.8	47115	3	US-09-949-016-12278	Sequence 12278, A	c 364	18.2	42.9	6450	3	US-09-949-016-16596	Sequence 16596, A
c 292	18.4	43.8	47122	3	US-09-949-016-16520	Sequence 16520, A	c 365	18.2	42.9	7430	3	US-08-956-171E-260	Sequence 260, App
c 293	18.4	43.8	83391	4	US-09-531-120-189	Sequence 189, App	c 366	18.2	42.9	7430	3	US-08-781-986A-260	Sequence 260, App
c 294	18.4	43.8	90336	4	US-09-531-120-195	Sequence 195, App	c 367	18.2	42.9	13086	3	US-08-956-171E-16	Sequence 16, Appl
c 295	18.4	43.8	95621	3	US-09-949-016-13237	Sequence 13237, A	c 368	18.2	42.9	13086	3	US-08-781-986A-16	Sequence 16, Appl
c 296	18.4	43.8	99629	3	US-09-596-002-37	Sequence 37, Appl	c 369	18.2	42.9	19181	3	US-09-949-016-15016	Sequence 15016, A
c 297	18.4	43.8	109250	3	US-09-949-016-12530	Sequence 12530, A	c 370	18.2	42.9	22206	3	US-09-949-016-13901	Sequence 13901, A
c 298	18.4	43.8	109251	3	US-09-949-016-17321	Sequence 17321, A	c 371	18.2	42.9	23155	3	US-09-949-016-12366	Sequence 12366, A
c 299	18.4	43.8	139049	3	US-09-949-016-17030	Sequence 17030, A	c 372	18.2	42.9	26760	3	US-09-949-016-15894	Sequence 15894, A
c 300	18.4	43.8	139257	3	US-09-920-671-11	Sequence 11, Appl	c 373	18.2	42.9	36090	3	US-09-949-016-17424	Sequence 17424, A
c 301	18.4	43.8	161124	3	US-09-949-016-11760	Sequence 11760, A	c 374	18.2	42.9	36123	3	US-09-949-016-12402	Sequence 12402, A
c 302	18.4	43.8	228851	3	US-09-949-016-13781	Sequence 13781, A	c 375	18.2	42.9	51629	3	US-09-949-016-12883	Sequence 12883, A
c 303	18.4	43.8	392000	3	US-10-027-983-11	Sequence 11, Appl	c 376	18.2	42.9	51631	3	US-09-949-016-15496	Sequence 15496, A
c 304	18.2	43.3	122	3	US-09-533-559-3329	Sequence 3329, Ap	c 377	18.2	42.9	66988	3	US-09-949-016-11942	Sequence 11942, A
c 305	18.2	43.3	375	3	US-09-252-891A-7801	Sequence 7801, Ap	c 378	18.2	42.9	66989	3	US-09-949-016-16063	Sequence 16063, A
c 306	18.2	43.3	601	3	US-09-949-016-170595	Sequence 170595, A	c 379	18.2	42.9	75378	3	US-09-949-016-17140	Sequence 17140, A
c 307	18.2	43.3	601	3	US-09-949-016-186297	Sequence 186297, A	c 380	18.2	42.9	86936	3	US-09-949-016-17314	Sequence 17314, A
c 308	18.2	43.3	601	3	US-09-949-016-201190	Sequence 201190, A	c 381	18.2	42.9	94133	3	US-09-949-016-11901	Sequence 11901, A
c 309	18.2	43.3	1272	3	US-09-328-925-82	Sequence 82, Appl	c 382	18.2	42.9	94133	3	US-09-949-016-12713	Sequence 12713, A
c 310	18.2	43.3	1272	3	US-09-328-925-83	Sequence 83, Appl	c 383	18.2	42.9	94135	3	US-09-949-016-15934	Sequence 15934, A
c 311	18.2	43.3	1273	3	US-09-328-925-84	Sequence 84, Appl	c 384	18.2	42.9	94135	3	US-09-949-016-15935	Sequence 15935, A
c 312	18.2	43.3	1273	3	US-09-328-925-76	Sequence 76, Appl	c 385	18.2	42.9	94135	3	US-09-949-016-15936	Sequence 15936, A
c 313	18.2	43.3	1273	3	US-09-328-925-77	Sequence 77, Appl	c 386	18.2	42.9	94135	3	US-09-949-016-15937	Sequence 15937, A
c 314	18.2	43.3	1273	3	US-09-328-925-78	Sequence 78, Appl	c 387	18.2	42.9	94755	3	US-09-949-016-11939	Sequence 11839, A
c 315	18.2	43.3	1273	3	US-09-328-925-79	Sequence 79, Appl	c 388	18.2	42.9	129415	3	US-09-949-016-16997	Sequence 16997, A



535	17.6	41.9	153642	3	US-09-949-016-12174	Sequence 12174, A	c 608	17.4	41.4	644	5	US-09-989-293A-282	Sequence 282, App
536	17.6	41.9	153643	3	US-09-949-016-15635	Sequence 15635, A	c 609	17.4	41.4	651	3	US-08-998-416-1087	Sequence 1087, A
c 537	17.6	41.9	172677	3	US-09-949-016-13444	Sequence 13444, A	c 610	17.4	41.4	700	3	US-09-735-271-1266	Sequence 266, App
538	17.6	41.9	173787	3	US-09-949-016-12542	Sequence 12542, A	c 611	17.4	41.4	700	3	US-09-735-271-1254	Sequence 1254, App
539	17.6	41.9	173791	3	US-09-949-016-17302	Sequence 17302, A	c 612	17.4	41.4	700	3	US-09-735-271-1255	Sequence 1255, App
c 540	17.6	41.9	198942	3	US-09-949-016-13209	Sequence 13209, A	c 613	17.4	41.4	742	3	US-08-702-705-1345	Sequence 1345, App
541	17.6	41.9	232547	3	US-09-949-016-16603	Sequence 16603, A	c 614	17.4	41.4	742	3	US-09-736-427-1345	Sequence 1345, App
542	17.6	41.9	239527	3	US-09-949-016-15980	Sequence 15980, A	c 615	17.4	41.4	742	3	US-09-614-124B-1345	Sequence 1345, App
c 543	17.6	41.9	278866	3	US-09-949-016-13922	Sequence 13922, A	c 616	17.4	41.4	742	3	US-09-671-325-1345	Sequence 1345, App
544	17.6	41.9	278866	3	US-09-949-016-13922	Sequence 13922, A	c 617	17.4	41.4	742	3	US-09-658-824-1345	Sequence 1345, App
c 545	17.6	41.9	278866	3	US-09-949-016-13924	Sequence 13924, A	c 618	17.4	41.4	742	3	US-10-017-754-1345	Sequence 1345, App
546	17.6	41.9	278866	3	US-09-949-016-13925	Sequence 13925, A	c 619	17.4	41.4	742	3	US-08-551-563-1345	Sequence 1345, App
c 547	17.6	41.9	278866	3	US-09-949-016-13926	Sequence 13926, A	c 620	17.4	41.4	1452	3	US-08-487-534A-50	Sequence 50, Appl
548	17.6	41.9	278866	3	US-09-949-016-14699	Sequence 14699, A	c 621	17.4	41.4	1452	3	US-09-527-431-50	Sequence 50, Appl
c 549	17.6	41.9	278866	3	US-09-949-016-14700	Sequence 14700, A	c 622	17.4	41.4	1452	3	US-09-446-861-50	Sequence 50, Appl
550	17.6	41.9	278866	3	US-09-949-016-14701	Sequence 14701, A	c 623	17.4	41.4	1452	3	US-09-107-532A-941	Sequence 941, App
c 551	17.6	41.9	278866	3	US-09-949-016-14702	Sequence 14702, A	c 624	17.4	41.4	1635	3	US-09-949-016-1439	Sequence 1439, App
552	17.6	41.9	278866	3	US-09-949-016-14703	Sequence 14703, A	c 625	17.4	41.4	1635	3	US-08-601-198-63	Sequence 63, Appl
c 553	17.6	41.9	300402	3	US-09-949-016-13632	Sequence 13632, A	c 626	17.4	41.4	2158	3	US-09-318-448-4	Sequence 4, Appl
554	17.6	41.9	390416	3	US-09-949-016-16923	Sequence 16923, A	c 627	17.4	41.4	2158	3	US-09-962-665-12	Sequence 12, Appl
c 555	17.6	41.9	524032	3	US-09-949-016-16928	Sequence 16928, A	c 628	17.4	41.4	2158	3	US-09-963-333-12	Sequence 12, Appl
556	17.6	41.9	524032	3	US-09-949-016-16928	Sequence 16928, A	c 629	17.4	41.4	2158	3	US-09-962-677-12	Sequence 12, Appl
c 557	17.6	41.9	524032	3	US-09-949-016-16930	Sequence 16930, A	c 630	17.4	41.4	2158	3	US-09-949-016-699	Sequence 699, App
558	17.6	41.9	524032	3	US-09-949-016-16931	Sequence 16931, A	c 631	17.4	41.4	2158	3	US-09-949-016-5835	Sequence 5835, App
c 559	17.6	41.9	529885	3	US-09-949-016-14340	Sequence 14340, A	c 632	17.4	41.4	2158	3	US-09-577-266-4	Sequence 4, Appl
560	17.6	41.9	529885	3	US-09-949-016-14341	Sequence 14341, A	c 633	17.4	41.4	22			



C 827	17	40.5	601	3	US-09-949-016-65408	Sequence 65408, A	C 900	17	40.5	3134	3	US-08-374-219B-7	Sequence 7, Appli
C 828	17	40.5	601	3	US-09-949-016-76585	Sequence 76585, A	C 901	17	40.5	3276	4	US-10-094-749-279	Sequence 279, App
C 829	17	40.5	601	3	US-09-949-016-77543	Sequence 77543, A	C 902	17	40.5	3646	3	US-09-710-279-4401	Sequence 4401, Ap
C 830	17	40.5	601	3	US-09-949-016-78852	Sequence 78852, A	C 903	17	40.5	3744	3	US-08-961-527-263	Sequence 263, App
C 831	17	40.5	601	3	US-09-949-016-78853	Sequence 78853, A	C 904	17	40.5	3795	3	US-10-002-344A-25	Sequence 25, Appli
C 832	17	40.5	601	3	US-09-949-016-107890	Sequence 107890, A	C 905	17	40.5	3921	2	US-08-567-375-3	Sequence 3, Appli
C 833	17	40.5	601	3	US-09-949-016-113399	Sequence 113399, A	C 906	17	40.5	3921	2	US-08-587-680A-3	Sequence 3, Appli
C 834	17	40.5	601	3	US-09-949-016-113400	Sequence 113400, A	C 907	17	40.5	3922	2	US-09-949-016-13969	Sequence 3969, Ap
C 835	17	40.5	601	3	US-09-949-016-123964	Sequence 123964, A	C 908	17	40.5	3923	3	US-09-949-016-844	Sequence 844, App
C 836	17	40.5	601	3	US-09-949-016-129176	Sequence 129176, A	C 909	17	40.5	5153	3	US-09-774-528-21	Sequence 21, Appl
C 837	17	40.5	601	3	US-09-949-016-129177	Sequence 129177, A	C 910	17	40.5	5153	3	US-10-120-988-21	Sequence 21, Appl
C 838	17	40.5	601	3	US-09-949-016-171946	Sequence 171946, A	C 911	17	40.5	5207	3	US-09-774-528-20	Sequence 20, Appl
C 839	17	40.5	601	3	US-09-949-016-180812	Sequence 180812, A	C 912	17	40.5	5207	3	US-10-120-988-20	Sequence 20, Appl
C 840	17	40.5	650	3	US-09-270-767-1808	Sequence 1808, Ap	C 913	17	40.5	5455	3	US-09-578-458-15	Sequence 15, Appl
C 841	17	40.5	650	3	US-09-270-767-17090	Sequence 17090, A	C 914	17	40.5	5965	3	US-09-362-336A-1	Sequence 1, Appli
C 842	17	40.5	669	3	US-08-956-171E-388	Sequence 388, App	C 915	17	40.5	5992	2	US-08-475-891A-3	Sequence 3, Appli
C 843	17	40.5	669	3	US-08-781-866A-388	Sequence 388, App	C 916	17	40.5	6014	5	US-09-984-429-368	Sequence 368, App
C 844	17	40.5	741	3	US-09-902-540-3629	Sequence 3629, Ap	C 917	17	40.5	6028	3	US-09-362-336A-3	Sequence 3, Appli
C 845	17	40.5	743	3	US-09-342-681C-11	Sequence 11, Appl	C 918	17	40.5	9038	3	US-09-949-016-13523	Sequence 13523, A
C 846	17	40.5	842	3	US-09-154-750A-72	Sequence 72, Appl	C 919	17	40.5	9210	3	US-09-712-363-100	Sequence 100, App
C 847	17	40.5	851	4	US-09-297-648-2987	Sequence 2987, Ap	C 920	17	40.5	9464	3	US-09-593-580B-1	Sequence 1, Appli
C 848	17	40.5	851	4	US-09-297-648-2988	Sequence 2988, Ap	C 921	17	40.5	9464	3	US-09-738-847-1	Sequence 1, Appli
C 849	17	40.5	879	3	US-09-976-594-987	Sequence 987, App	C 922	17	40.5	9464	3	US-09-738-847-1	Sequence 1, Appli
C 850	17	40.5	913	3	US-09-533-559-3441	Sequence 3441, Ap	C 923	17	40.5	11378	3	US-09-949-016-13815	Sequence 13815, A
C 851	17	40.5	999	3	US-09-252-991A-14199	Sequence 14199, A	C 924	17	40.5	15000	3	US-09-719-554-65	Sequence 65, Appl
C 852	17	40.5	1052	3	US-09-270-767-27177	Sequence 27177, A	C 925	17	40.5	15206	3	US-09-949-016-13585	Sequence 13585, A
C 853	17	40.5	1092	3	US-09-252-991A-16279	Sequence 16279, A	C 926	17	40.5	15206	3	US-09-949-016-13586	Sequence 13586, A
C 854	17	40.5	1141	3	US-09-617-174C-1	Sequence 1, Appli	C 927	17	40.5	15598	3	US-08-956-171E-82	Sequence 82, Appl
C 855	17	40.5	1179	3	US-09-491-577-81	Sequence 81, Appl	C 928	17	40.5	15598	3	US-08-781-896A-82	Sequence 82, Appl
C 856	17	40.5	1194	3	US-09-134-000C-2759	Sequence 2759, Ap	C 929	17	40.5	16442	3	US-08-781-896A-208	Sequence 208, App
C 857	17	40.5	1305	3	US-09-540-236-933	Sequence 933, App	C 930	17	40.5	16442	3	US-09-618-166-208	Sequence 208, App
C 858	17	40.5	1438	3	US-09-187-331-4	Sequence 4, Appli	C 931	17	40.5	17198	3	US-09-949-016-13776	Sequence 13776, A
C 859	17	40.5	1438	3	US-09-470-966-4	Sequence 4, Appli	C 932	17	40.5	17318	3	US-09-949-016-13817	Sequence 13817, A
C 860	17	40.5	1438	3	US-08-928-069-5	Sequence 5, Appli	C 933	17	40.5	17591	3	US-09-949-002-658	Sequence 658, App
C 861	17	40.5	1438	3	US-08-828-683A-5	Sequence 5, Appli	C 934	17	40.5	17591	3	US-09-949-002-828	Sequence 828, App
C 862	17	40.5	1438	3	US-09-339-159B-3	Sequence 3, Appli	C 935	17	40.5	18082	3	US-09-949-016-12586	Sequence 12586, A
C 863	17	40.5	1470	3	US-09-339-159B-1	Sequence 1, Appli	C 936	17	40.5	18082	3	US-09-949-016-15711	Sequence 15711, A
C 864	17	40.5	1579	4	US-10-036-041-8	Sequence 8, Appli	C 937	17	40.5	18474	3	US-09-949-016-17559	Sequence 17559, A
C 865	17	40.5	1581	3	US-09-134-000C-3099	Sequence 3099, Ap	C 938	17	40.5	18848	3	US-09-902-540-1174	Sequence 1174, Ap
C 866	17	40.5	1791	3	US-09-270-767-11575	Sequence 11575, A	C 939	17	40.5	25190	3	US-09-949-016-15906	Sequence 15906, A
C 867	17	40.5	1959	3	US-09-252-991A-14026	Sequence 14026, A	C 940	17	40.5	27295	3	US-09-949-016-16605	Sequence 16605, A
C 868	17	40.5	2001	3	US-09-799-451-695	Sequence 695, App	C 941	17	40.5	29272	3	US-09-902-540-1217	Sequence 1217, Ap
C 869	17	40.5	2004	4	US-10-094-749-1462	Sequence 1462, Ap	C 942	17	40.5	29954	3	US-09-949-016-13808	Sequence 13808, A
C 870	17	40.5	2174	3	US-09-634-238-54	Sequence 54, Appl	C 943	17	40.5	31040	3	US-09-949-016-12383	Sequence 12383, A
C 871	17	40.5	2184	3	US-09-540-236-1103	Sequence 1103, Ap	C 944	17	40.5	31063	3	US-09-596-002-20	Sequence 20, Appl
C 872	17	40.5	2197	3	US-08-985-950-13	Sequence 13, Appl	C 945	17	40.5	31166	3	US-09-949-016-13464	Sequence 13464, A
C 873	17	40.5	2197	3	US-09-546-049-13	Sequence 13, Appl	C 946	17	40.5	36090	3	US-09-949-016-17424	Sequence 17424, A
C 874	17	40.5	2200	3	US-08-985-950-19	Sequence 19, Appl	C 947	17	40.5	36123	3	US-09-949-016-12402	Sequence 12402, A
C 875	17	40.5	2200	3	US-09-546-049-19	Sequence 19, Appl	C 948	17	40.5	45225	3	US-09-949-016-12428	Sequence 12428, A
C 876	17	40.5	2211	3	US-09-107-532A-2353	Sequence 2353, Ap	C 949	17	40.5	45226	3	US-09-949-016-13654	Sequence 13654, A
C 877	17	40.5	2211	3	US-09-310-463-9	Sequence 9, Appli	C 950	17	40.5	45842	3	US-09-949-016-16914	Sequence 16914, A
C 878	17	40.5	2221	3	US-08-842-248A-9	Sequence 9, Appli	C 951	17	40.5	51508	3	US-09-949-016-16681	Sequence 16681, A
C 879	17	40.5	2221	5	US-10-143-618-9	Sequence 9, Appli	C 952	17	40.5	56523	3	US-09-949-016-14287	Sequence 14287, A
C 880	17	40.5	2271	3	US-08-985-950-15	Sequence 15, Appl	C 953	17	40.5	58361	3	US-09-949-016-16755	Sequence 16755, A
C 881	17	40.5	2271	3	US-09-546-049-15	Sequence 15, Appl	C 954	17	40.5	58361	3	US-09-949-016-16756	Sequence 16756, A
C 882	17	40.5	2271	3	US-09-252-991A-14238	Sequence 14238, A	C 955	17	40.5	59065	3	US-09-813-817-3	Sequence 3, Appli
C 883	17	40.5	2291	5	US-09-974-300-6330	Sequence 6330, Ap	C 956	17	40.5	59065	3	US-09-978-197-3	Sequence 3, Appli
C 884	17	40.5	2388	3	US-08-985-950-17	Sequence 17, Appl	C 957	17	40.5	59065	3	US-10-135-696-3	Sequence 3, Appli
C 885	17	40.5	2388	3	US-09-546-049-17	Sequence 17, Appl	C 958	17	40.5	59065	3	US-10-820-230-3	Sequence 3, Appli
C 886	17	40.5	2462	3	US-10-104-047-78	Sequence 78, Appl	C 959	17	40.5	59519	3	US-09-949-016-13504	Sequence 13504, A
C 887	17	40.5	2481	3	US-08-899-578-1	Sequence 1, Appli	C 960	17	40.5	60489	3	US-09-949-016-16287	Sequence 16287, A
C 888	17	40.5	2574	3	US-09-248-796A-6419	Sequence 6419, Ap	C 961	17	40.5	61042	3	US-09-949-016-13224	Sequence 13224, A
C 889	17	40.5	2576	4	US-10-094-749-1045	Sequence 1045, Ap	C 962	17	40.5	61735	3	US-09-949-016-12064	Sequence 12064, A
C 890	17	40.5	2724	3	US-10-104-047-1127	Sequence 1127, Ap	C 963	17	40.5	67620	3	US-09-949-016-16939	Sequence 16939, A
C 891	17	40.5	2860	2	US-08-462-484-7	Sequence 7, Appli	C 964	17	40.5	67911	3	US-09-949-016-16979	Sequence 16979, A
C 892	17	40.5	2860	2	US-08-441-147-7	Sequence 7, Appli	C 965	17	40.5	71815	3	US-09-949-016-12501	Sequence 12501, A
C 893	17	40.5	2860	7	PCT-US95-07536-7	Sequence 7, Appli	C 966	17	40.5	74527	3	US-09-949-016-12339	Sequence 12339, A
C 894	17	40.5	2909	3	US-10-104-047-614	Sequence 614, App	C 967	17	40.5	74528	3	US-09-949-016-13275	Sequence 13275, A
C 895	17	40.5	3001	3	US-09-539-333D-181	Sequence 181, App	C 968	17	40.5	82596	4	US-09-531-120-207	Sequence 207, App
C 896	17	40.5	3001	3	US-09-539-333D-197	Sequence 197, App	C 969	17	40.5	86439	3	US-09-949-016-11945	Sequence 11945, A
C 897	17	40.5	3078	3	US-09-623-551-28	Sequence 28, Appl	C 970	17	40.5	86440	3	US-09-949-016-16990	Sequence 16990, A
C 898	17	40.5	3092	3	US-09-566-921-121	Sequence 121, App	C 971	17	40.5	86980	3	US-09-949-016-15344	Sequence 15344, A
C 899	17	40.5	3134	2	US-07-865-662F-7	Sequence 7, Appli	C 972	17	40.5	92584	4	US-09-531-120-194	Sequence 194, App

973 17 40.5 94873 3 US-09-949-016-14277 Sequence 14277, A  
974 17 40.5 94905 4 US-09-531-120-208 Sequence 208, App  
975 17 40.5 98864 3 US-09-949-016-15403 Sequence 15403, A  
976 17 40.5 99304 3 US-09-949-016-15440 Sequence 15440, A  
977 17 40.5 103750 3 US-09-949-016-13319 Sequence 13319, A  
978 17 40.5 105055 3 US-09-949-016-14001 Sequence 14001, A  
979 17 40.5 105733 3 US-09-949-016-13080 Sequence 13080, A  
980 17 40.5 118067 3 US-09-497-855A-32 Sequence 32, Appl  
981 17 40.5 124110 3 US-09-949-016-13353 Sequence 13353, A  
982 17 40.5 134499 4 US-09-531-120-192 Sequence 192, App  
983 17 40.5 135010 3 US-09-949-016-17234 Sequence 17234, A  
984 17 40.5 139250 3 US-09-949-016-17398 Sequence 17398, A  
985 17 40.5 139577 3 US-09-949-016-12879 Sequence 12879, A  
986 17 40.5 141589 5 US-09-543-679A-2480 Sequence 2480, Ap  
987 17 40.5 141589 5 US-09-543-679A-2694 Sequence 2694, Ap  
988 17 40.5 141589 5 US-09-543-679A-2719 Sequence 2719, Ap  
989 17 40.5 145241 3 US-09-949-016-17394 Sequence 17394, A  
990 17 40.5 145941 3 US-09-949-016-17395 Sequence 17395, A  
991 17 40.5 146982 5 US-09-543-679A-3009 Sequence 3009, Ap  
992 17 40.5 148794 3 US-09-949-016-12751 Sequence 12751, A  
993 17 40.5 149543 3 US-09-949-016-15947 Sequence 15947, A  
994 17 40.5 150597 3 US-09-949-016-15379 Sequence 15379, A  
995 17 40.5 151828 4 US-09-531-120-197 Sequence 197, App  
996 17 40.5 152524 3 US-09-949-016-12683 Sequence 12683, A  
997 17 40.5 152524 3 US-09-949-016-13194 Sequence 13194, A  
998 17 40.5 168174 3 US-10-071-411A-63 Sequence 63, Appl  
999 17 40.5 168273 3 US-10-071-411A-2 Sequence 2, Appli  
ci000 17 40.5 177293 3 US-09-949-016-16513 Sequence 16513, A

## ALIGNMENTS

RESULT 1  
US-09-949-016-29173  
; Sequence 29173, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 29173  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-29173

Query Match 100.0%; Score 42; DB 3; Length 601;  
Best Local Similarity 100.0%; Pred. No. 7.8e-09;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
Db 14 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 55

RESULT 2  
US-09-949-016-126957  
; Sequence 126957, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 126957  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-126957

Query Match 100.0%; Score 42; DB 3; Length 601;  
Best Local Similarity 100.0%; Pred. No. 7.8e-09;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
Db 14 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 55

RESULT 3  
US-09-188-930-249  
; Sequence 249, Application US/09188930A  
; Patent No. 6150502  
; GENERAL INFORMATION:  
; APPLICANT: Watson, James D.  
; APPLICANT: Strachan, Lorna  
; APPLICANT: Sleeman, Matthew  
; APPLICANT: Chrust, Rene  
; APPLICANT: Murison, James Greg  
; TITLE OF INVENTION: Compositions Isolated From Skin Cells  
; AND METHODS FOR THEIR USE  
; FILE REFERENCE: 11000.1011c1  
; CURRENT APPLICATION NUMBER: US/09/188,930A  
; CURRENT FILING DATE: 1998-11-09  
; NUMBER OF SEQ ID NOS: 348  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 249  
; LENGTH: 1212  
; TYPE: DNA  
; ORGANISM: Human  
US-09-188-930-249

Query Match 100.0%; Score 42; DB 3; Length 1212;  
Best Local Similarity 100.0%; Pred. No. 9.4e-09;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
Db 990 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1031

RESULT 4  
US-09-312-283C-249  
; Sequence 249, Application US/09312283C  
; Patent No. 6573095  
; GENERAL INFORMATION:  
; APPLICANT: Watson, James D.  
; APPLICANT: Strachan, Lorna  
; APPLICANT: Sleeman, Matthew  
; APPLICANT: Chrust, Rene  
; APPLICANT: Murison, James G.  
; APPLICANT: Kumble, Krishanand D.  
; TITLE OF INVENTION: Compositions Isolated from Skin Cells  
; AND METHODS FOR THEIR USE

; FILE REFERENCE: 11000.1011c2  
; CURRENT APPLICATION NUMBER: US/09/312,283C  
; CURRENT FILING DATE: 1999-05-14  
; NUMBER OF SEQ ID NOS: 425  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 249  
; LENGTH: 1212  
; TYPE: DNA  
; ORGANISM: Mouse  
US-09-312-283C-249

Query Match 100.0%; Score 42; DB 3; Length 1212;  
Best Local Similarity 100.0%; Pred. No. 9.4e-09;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGGCGAGCCCTTACAAGACTATGAAGTAAAG 42  
|||||  
Db 990 GCCACGCTTGGCGAGCCCTTACAAGACTATGAAGTAAAG 1031  
|||||

## RESULT 5

US-09-495-050A-223

; Sequence 223, Application US/09495050A

; Patent No. 6492505

; GENERAL INFORMATION:

; APPLICANT: Roopa, Reddy

; APPLICANT: Guegler, Karl, J.

; APPLICANT: Au-Young, Janice

; TITLE OF INVENTION: COMPOSITION FOR DETECTION OF GENES ENCODING MEMBRANE-ASSOCIATED H

; FILE REFERENCE: PA-0013 US

; CURRENT APPLICATION NUMBER: US/09/495,050A

; CURRENT FILING DATE: 2000-01-31

; PRIOR APPLICATION NUMBER: 60/118,318

; PRIOR FILING DATE: February 1, 1999

; NUMBER OF SEQ ID NOS: 305

; SOFTWARE: PERL Program

; SEQ ID NO 223

; LENGTH: 1228

; TYPE: DNA

; ORGANISM: Homo sapiens

; NAME/KEY: misc feature

; OTHER INFORMATION: Incyte ID No. 6492505 2681738CT1

US-09-495-050A-223

Query Match 100.0%; Score 42; DB 3; Length 1228;  
Best Local Similarity 100.0%; Pred. No. 9.5e-09;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGGCGAGCCCTTACAAGACTATGAAGTAAAG 42  
|||||  
Db 541 GCCACGCTTGGCGAGCCCTTACAAGACTATGAAGTAAAG 582  
|||||

## RESULT 6

US-09-949-016-3613

; Sequence 3613, Application US/09949016

; Patent No. 6812339

; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

; FILE REFERENCE: CL001307

; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 2000-04-14

; PRIOR APPLICATION NUMBER: 60/241,755

; PRIOR FILING DATE: 2000-10-20

; PRIOR APPLICATION NUMBER: 60/237,768

; PRIOR FILING DATE: 2000-10-03

; PRIOR APPLICATION NUMBER: 60/231,498

; PRIOR FILING DATE: 2000-09-08

; NUMBER OF SEQ ID NOS: 207012

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 3613  
; LENGTH: 2289  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-3613

Query Match 100.0%; Score 42; DB 3; Length 2289;  
Best Local Similarity 100.0%; Pred. No. 1.1e-08;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
|||||  
Db 1117 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 1158  
|||||

## RESULT 7

US-09-949-016-572

; Sequence 572, Application US/09949016

; Patent No. 6812339

; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

; FILE REFERENCE: CL001307

; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 2000-04-14

; PRIOR APPLICATION NUMBER: 60/241,755

; PRIOR FILING DATE: 2000-10-20

; PRIOR APPLICATION NUMBER: 60/237,768

; PRIOR FILING DATE: 2000-10-03

; PRIOR APPLICATION NUMBER: 60/231,498

; PRIOR FILING DATE: 2000-09-08

; NUMBER OF SEQ ID NOS: 207012

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 572

; LENGTH: 2456

; TYPE: DNA

; ORGANISM: Human

US-09-949-016-572

Query Match 100.0%; Score 42; DB 3; Length 2456;  
Best Local Similarity 100.0%; Pred. No. 1.1e-08;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
|||||  
Db 1271 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 1312  
|||||

## RESULT 8

US-09-949-016-15355

; Sequence 15355, Application US/09949016

; Patent No. 6812339

; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

; FILE REFERENCE: CL001307

; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 2000-04-14

; PRIOR APPLICATION NUMBER: 60/241,755

; PRIOR FILING DATE: 2000-10-20

; PRIOR APPLICATION NUMBER: 60/237,768

; PRIOR FILING DATE: 2000-10-03

; PRIOR APPLICATION NUMBER: 60/231,498

; PRIOR FILING DATE: 2000-09-08

; NUMBER OF SEQ ID NOS: 207012

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 15355

; LENGTH: 26086

; TYPE: DNA

; ORGANISM: Human

US-09-949-016-15355

```
Query Match      100.0%; Score 42; DB 3; Length 26086;
Best Local Similarity 100.0%; Pred. No. 2.2e-08;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACGCTTGCCGCGCCCTTACAAGACTATGAAGTAAAG 42
DB 22913 GCCACGCTTGCCGCGCCCTTACAAGACTATGAAGTAAAG 22954

RESULT 9
US-09-949-016-12314
; Sequence 12314, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12314
; LENGTH: 26238
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12314

Query Match      100.0%; Score 42; DB 3; Length 26238;
Best Local Similarity 100.0%; Pred. No. 2.2e-08;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACGCTTGCCGCGCCCTTACAAGACTATGAAGTAAAG 42
DB 23067 GCCACGCTTGCCGCGCCCTTACAAGACTATGAAGTAAAG 23108

RESULT 10
US-09-555-704-3
; Sequence 3, Application US/09555704
; Patent No. 6841374
; GENERAL INFORMATION:
; APPLICANT: The Research Foundation for Microbial diseases of Osaka University
; TITLE OF INVENTION: Japanese Encephalitis
; FILE REFERENCE: P98AF308-2
; CURRENT APPLICATION NUMBER: US/09/555,704
; CURRENT FILING DATE: 2000-06-02
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1500
; TYPE: DNA
; ORGANISM: Japanese encephalitis virus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1500)
US-09-555-704-3

Query Match      54.3%; Score 22.8; DB 3; Length 1500;
Best Local Similarity 79.4%; Pred. No. 3.6;
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 CACGCTTGCCGCGCCCTTACAAGACTATGAAG 37
DB 1203 CACGCTGGCGAAGCCCTTTCAACGACTTTGAAG 1236
```

```
RESULT 11
US-08-224-391-52
; Sequence 52, Application US/08224391
; Patent No. 5744140
; GENERAL INFORMATION:
; APPLICANT: Paoletti, Enzo
; APPLICANT: Pincus, Steven E.
; TITLE OF INVENTION: FLAVIVIRUS RECOMBINANT POXVIRUS VACCINE
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford
; ADDRESSEE: c/o William S. Frommer
; STREET: 530 Fifth Avenue
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/224,391
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/729,800
; FILING DATE: 17-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer, William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454310-2340
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4512 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-224-391-52

Query Match      54.3%; Score 22.8; DB 2; Length 4512;
Best Local Similarity 79.4%; Pred. No. 4.9;
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 CACGCTTGCCGCGCCCTTACAAGACTATGAAG 37
DB 2085 CACGCTGGCGAAGCCCTTTCAACGACTTTGAAG 2118

RESULT 12
US-08-484-304-52
; Sequence 52, Application US/08484304
; Patent No. 5744141
; GENERAL INFORMATION:
; APPLICANT: Paoletti, Enzo
; APPLICANT: Pincus, Steven E.
; TITLE OF INVENTION: FLAVIVIRUS RECOMBINANT POXVIRUS VACCINE
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford
; ADDRESSEE: c/o William S. Frommer
; STREET: 530 Fifth Avenue
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```



; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/484,304  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/224,391  
; FILING DATE:  
; APPLICATION NUMBER: US 07/729,800  
; FILING DATE: 17-JUL-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Frommer, William S.  
; REGISTRATION NUMBER: 25,506  
; REFERENCE/DOCKET NUMBER: 454310-2340  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 840-3333  
; TELEFAX: (212) 840-0712  
; INFORMATION FOR SEQ ID NO: 52:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4512 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-484-304-52

Query Match 54.3%; Score 22.8; DB 2; Length 4512;  
Best Local Similarity 79.4%; Pred. No. 4.9;  
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 4 CACGCTGGCGGAGCCCTTACAAAGACTATGAAG 37  
Db 2085 CACGCTGGCGGAGCCCTTTCAACGACTTTGAAG 2118

## RESULT 13

US-09-949-016-14107  
; Sequence 14107, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14107  
; LENGTH: 19319  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-14107

Query Match 51.9%; Score 21.8; DB 3; Length 19319;  
Best Local Similarity 70.7%; Pred. No. 21;  
Matches 29; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 2 CCCCAGCTTGGCGGAGCCCTTACAAAGACTATGAAGTAAAG 42  
Db 13349 CCCCAGCTTGAAGAGGTCTTATATAAACTTTCAAGTAAAG 13389

## RESULT 14

US-09-513-999C-20939/c  
; Sequence 20939, Application US/09513999C

; Patent No. 6783961  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Duclert, A.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.  
; Patent No. 6783961  
; FILE REFERENCE: 59 US2 REG  
; CURRENT APPLICATION NUMBER: US/09/513,999C  
; CURRENT FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/122,487  
; PRIOR FILING DATE: 1999-02-26  
; NUMBER OF SEQ ID NOS: 36681  
; SOFTWARE: Patent.pm  
; SEQ ID NO 20939  
; LENGTH: 452  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 9  
; OTHER INFORMATION: w=a or t  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 10  
; OTHER INFORMATION: k=g or t  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 11  
; OTHER INFORMATION: k=g or t  
US-09-513-999C-20939

Query Match 51.4%; Score 21.6; DB 3; Length 452;  
Best Local Similarity 75.0%; Pred. No. 9;  
Matches 27; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCGGAGCCCTTACAAAGACTATGAA 36  
Db 400 GTCCAGGCTTGGTCCAGCCCTTAGAAGGGCTAGCAA 365

## RESULT 15

US-09-949-016-177622  
; Sequence 177622, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 177622  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-177622

Query Match 51.4%; Score 21.6; DB 3; Length 601;  
Best Local Similarity 75.0%; Pred. No. 9.7;  
Matches 27; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCGGAGCCCTTACAAAGACTATGAA 36  
Db 169 GTCCAGGCTTGGTCCAGCCCTTAGAAGGGCTAGCAA 204

```

; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16819
; LENGTH: 76399
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)....(76399)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16819

Query Match 51.4%; Score 21.6; DB 3; Length 76399;
Best Local Similarity 75.0%; Pred. No. 37;
Matches 27; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAA 36
Db 45406 GTCCAGGCTTGGTCCAGCCCTTAGAAGGGCTAGCAA 45371

RESULT 19
US-09-513-999C-32761
; Sequence 32761, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59 US2 REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 32761
; LENGTH: 407
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 384
; OTHER INFORMATION: r=a or g
US-09-513-999C-32761

Query Match 51.0%; Score 21.4; DB 3; Length 407;
Best Local Similarity 80.6%; Pred. No. 11;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 11 GGCCGAGCCCTTACAAAGACTATGAAGTAAA 41
Db 10 GGCATGGCCCTTCACAACTGACTATGAAGTAAA 40

RESULT 20
US-09-949-016-12756/c
; Sequence 12756, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 177623
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-177623

Query Match 51.4%; Score 21.6; DB 3; Length 601;
Best Local Similarity 75.0%; Pred. No. 9.7;
Matches 27; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAA 36
Db 467 GTCCAGGCTTGGTCCAGCCCTTAGAAGGGCTAGCAA 502

RESULT 17
US-09-851-896-3/c
; Sequence 3, Application US/09851896
; Patent No. 6410325
; GENERAL INFORMATION:
; APPLICANT: C. Frank Bennett
; APPLICANT: Susan M. Freier
; APPLICANT: Andrew T. Watt
; TITLE OF INVENTION: ANTISENSE MODULATION OF PHOSPHOLIPASE A2, GROUP VI (CA2+-INDEPENDENT)
; FILE REFERENCE: RTS-0220
; CURRENT APPLICATION NUMBER: US/09/851,896
; CURRENT FILING DATE: 2001-05-08
; NUMBER OF SEQ ID NOS: 89
; SEQ ID NO 3
; LENGTH: 70000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-851-896-3

Query Match 51.4%; Score 21.6; DB 3; Length 70000;
Best Local Similarity 75.0%; Pred. No. 36;
Matches 27; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAA 36
Db 42352 GTCCAGGCTTGGTCCAGCCCTTAGAAGGGCTAGCAA 42317

RESULT 18
US-09-949-016-16819/c
; Sequence 16819, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
```



```
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1950
; LENGTH: 2772
; TYPE: DNA
; ORGANISM: Human
; ORGANISM: Human
US-09-949-016-1950

Query Match
Best Local Similarity 49.0%; Score 20.6; DB 3; Length 2772;
Matches 26; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 8 CTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42
Db 1811 COTGGCGAGCGATCACAAATTACATGTCAGTAAAG 1845

RESULT 26
US-09-949-016-13692
; Sequence 13692, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13692
; LENGTH: 36895
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13692

Query Match
Best Local Similarity 49.0%; Score 20.6; DB 3; Length 36895;
Matches 26; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 8 CTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42
Db 33934 COTGGCGAGCGATCACAAATTACATGTCAGTAAAG 33968

RESULT 27
US-09-702-705-823/c
; Sequence 823, Application US/09702705
; Patent No. 6504010
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedwick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.478C14
; CURRENT APPLICATION NUMBER: US/09/702,705
; CURRENT FILING DATE: 2000-10-30
; NUMBER OF SEQ ID NOS: 1833
```

---

```
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13276
; LENGTH: 100877
; TYPE: DNA
; ORGANISM: Human
; ORGANISM: Human
; NAME/KEY: misc.feature
; LOCATION: (1)...(100877)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13276

Query Match
Best Local Similarity 49.5%; Score 20.8; DB 3; Length 100877;
Matches 28; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1 GCCACGCTTGCCGAGCCCTTACAAAGACTATGAAGTAA 40
Db 18788 GACCACGCTAGACAGTGCAGGAACAAAACATAATGTAA 18749

RESULT 24
US-09-949-016-17074/c
; Sequence 17074, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17074
; LENGTH: 140725
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-17074

Query Match
Best Local Similarity 49.5%; Score 20.8; DB 3; Length 140725;
Matches 28; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 3 CCACGCTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42
Db 29158 CCACGCCAGCCCTAGCCCTTTCTAAGACCTGTAGTAAAG 29119

RESULT 25
US-09-949-016-1950
; Sequence 1950, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
```

```
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 823
; LENGTH: 264
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(264)
; OTHER INFORMATION: n=A,T,C or G
US-09-702-705-823

Query Match      48.6%; Score 20.4; DB 3; Length 264;
Best Local Similarity 64.3%; Pred. No. 26;
Matches 27; Conservative 0; Mismatches 0; Gaps 0;

Qy 1 GCCCAGCTTGGCGGAGCCCTTACAAAGACTATGAAGTAAAG 42
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 192 GCCCGGTNTGGCCNANCGCTGGCAAAGGCCCCAGNAGGAAAG 151

RESULT 28
US-09-736-457-823/c
; Sequence 823, Application US/09736457
; Patent No. 6509448
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.478C15
; CURRENT APPLICATION NUMBER: US/09/736.457
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 1864
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 823
; LENGTH: 264
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(264)
; OTHER INFORMATION: n=A,T,C or G
US-09-736-457-823

Query Match      48.6%; Score 20.4; DB 3; Length 264;
Best Local Similarity 64.3%; Pred. No. 26;
Matches 27; Conservative 0; Mismatches 0; Gaps 0;

Qy 1 GCCCAGCTTGGCGGAGCCCTTACAAAGACTATGAAGTAAAG 42
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 192 GCCCGGTNTGGCCNANCGCTGGCAAAGGCCCCAGNAGGAAAG 151

RESULT 29
US-09-614-124B-823/c
; Sequence 823, Application US/09614124B
; Patent No. 6630574
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
```

```
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; FILE REFERENCE: 210121.478C9
; CURRENT APPLICATION NUMBER: US/09/614.124B
; CURRENT FILING DATE: 2001-07-11
; NUMBER OF SEQ ID NOS: 1668
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 823
; LENGTH: 264
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(264)
; OTHER INFORMATION: n=A,T,C or G
US-09-614-124B-823

Query Match      48.6%; Score 20.4; DB 3; Length 264;
Best Local Similarity 64.3%; Pred. No. 26;
Matches 27; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCGGAGCCCTTACAAAGACTATGAAGTAAAG 42
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 192 GCCCGGTNTGGCCNANCGCTGGCAAAGGCCCCAGNAGGAAAG 151

RESULT 30
US-09-671-325-823/c
; Sequence 823, Application US/09671325
; Patent No. 6667154
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.478C12
; CURRENT APPLICATION NUMBER: US/09/671.325
; CURRENT FILING DATE: 2000-09-26
; NUMBER OF SEQ ID NOS: 1825
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 823
; LENGTH: 264
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(264)
; OTHER INFORMATION: n=A,T,C or G
US-09-671-325-823

Query Match      48.6%; Score 20.4; DB 3; Length 264;
Best Local Similarity 64.3%; Pred. No. 26;
Matches 27; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCGGAGCCCTTACAAAGACTATGAAGTAAAG 42
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 192 GCCCGGTNTGGCCNANCGCTGGCAAAGGCCCCAGNAGGAAAG 151

RESULT 31
US-09-589-184-823/c
; Sequence 823, Application US/09589184
; Patent No. 6686447
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
```

APPLICANT: Bangur, Chaitanya S.  
APPLICANT: Lodes, Michael A.  
APPLICANT: Fanger, Gary  
APPLICANT: Vedvick, Tom  
APPLICANT: Carter, Darrick  
APPLICANT: Retter, Marc  
APPLICANT: Mannion, Jane  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND  
TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER  
FILE REFERENCE: 210121.478C8  
CURRENT APPLICATION NUMBER: US/09/589,184  
CURRENT FILING DATE: 2000-06-05  
NUMBER OF SEQ ID NOS: 827  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 823  
LENGTH: 264  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(264)  
OTHER INFORMATION: n=A,T,C or G  
US-09-589-184-823

Query Match 48.6%; Score 20.4; DB 3; Length 264;  
Best Local Similarity 64.3%; Pred. No. 26;  
Matches 27; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 GCCACGCTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
|||||  
Db 192 GCCCCGGTGGCCNANCGCTGCGAAAGGCCAGNAGGAAAG 151

## RESULT 32

US-09-658-824-823/c  
Sequence 823, Application US/09658824  
Patent No. 6746846

GENERAL INFORMATION:  
APPLICANT: Wang, Tongtong  
APPLICANT: Bangur, Chaitanya S.  
APPLICANT: Lodes, Michael A.  
APPLICANT: Fanger, Gary  
APPLICANT: Vedvick, Tom  
APPLICANT: Carter, Darrick  
APPLICANT: Retter, Marc  
APPLICANT: Mannion, Jane  
APPLICANT: Fan, Liqun  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND  
TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER  
FILE REFERENCE: 210121.478C11  
CURRENT APPLICATION NUMBER: US/09/658,824  
CURRENT FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 1788  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 823  
LENGTH: 264  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(264)  
OTHER INFORMATION: n=A,T,C or G  
US-09-658-824-823

Query Match 48.6%; Score 20.4; DB 3; Length 264;  
Best Local Similarity 64.3%; Pred. No. 26;  
Matches 27; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 GCCACGCTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
|||||  
Db 192 GCCCCGGTGGCCNANCGCTGCGAAAGGCCAGNAGGAAAG 151

## RESULT 33

US-10-017-754-823/c  
Sequence 823, Application US/10017754  
Patent No. 6858204

GENERAL INFORMATION:  
APPLICANT: Henderson, Robert A.  
APPLICANT: Wang, Tongtong  
APPLICANT: Watanabe, Yoshihiro  
APPLICANT: Johnson, Jeffrey C.  
APPLICANT: Retter, Marc W.  
APPLICANT: Marnerakis, Margarita  
APPLICANT: Carter, Darrick  
APPLICANT: Fanger, Gary R.  
APPLICANT: Vedvick, Thomas S.  
APPLICANT: Bangur, Chaitanya S.  
APPLICANT: McNabb, Andria  
TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
FILE REFERENCE: 210121.478C18  
CURRENT APPLICATION NUMBER: US/10/017,754  
CURRENT FILING DATE: 2001-10-29  
NUMBER OF SEQ ID NOS: 2004  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 823  
LENGTH: 264  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1, 4, 7, 15, 17, 35, 38, 44, 53, 90, 105, 108, 115, 117,  
LOCATION: 121, 126, 128, 158, 176, 178, 184, 201, 221, 227, 229, 233,  
LOCATION: 239, 250  
OTHER INFORMATION: n = A,T,C or G  
US-10-017-754-823

Query Match 48.6%; Score 20.4; DB 3; Length 264;  
Best Local Similarity 64.3%; Pred. No. 26;  
Matches 27; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 GCCACGCTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
|||||  
Db 192 GCCCCGGTGGCCNANCGCTGCGAAAGGCCAGNAGGAAAG 151

## RESULT 34

US-09-651-563-823/c  
Sequence 823, Application US/09651563  
Patent No. 6914132

GENERAL INFORMATION:  
APPLICANT: Wang, Tongtong  
APPLICANT: Bangur, Chaitanya S.  
APPLICANT: Lodes, Michael A.  
APPLICANT: Fanger, Gary  
APPLICANT: Vedvick, Tom  
APPLICANT: Carter, Darrick  
APPLICANT: Retter, Marc  
APPLICANT: Mannion, Jane  
APPLICANT: Fan, Liqun  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND  
TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER  
FILE REFERENCE: 210121.478C10  
CURRENT APPLICATION NUMBER: US/09/651,563  
CURRENT FILING DATE: 2000-08-29  
NUMBER OF SEQ ID NOS: 1679  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 823  
LENGTH: 264  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(264)  
OTHER INFORMATION: n=A,T,C or G



Query Match	48.6%; Score 20.4; DB 4; Length 2523;
Best Local Similarity	71.1%; Pred. No. 49;
Matches	27; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
QY	1 GCCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGT 38 
Db	185 GCCCTTGCGGCCTTGGCCCCCTTACAGACTATGTGGT 222 
RESULT 39	
US-10-104-047-349	
; Sequence 349, Application US/10104047	
; Patent No. 6943241	
; GENERAL INFORMATION:	
; APPLICANT: HELIX RESEARCH INSTITUTE	
; TITLE OF INVENTION: No. 6943241el full length cDNA	
; FILE REFERENCE: H1-A0105	
; CURRENT APPLICATION NUMBER: US/10/104,047	
; CURRENT FILING DATE: 2002-03-25	
; PRIOR APPLICATION NUMBER:	
; PRIOR FILING DATE:	
; NUMBER OF SEQ ID NOS: 4096	
; SOFTWARE: PatentIn Ver. 2.1	
; SEQ ID NO 349	
; LENGTH: 2648	
; TYPE: DNA	
; ORGANISM: Homo sapiens	
US-10-104-047-349	
Query Match	48.6%; Score 20.4; DB 3; Length 2648;
Best Local Similarity	71.1%; Pred. No. 50;
Matches	27; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
QY	1 GCCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGT 38 
Db	659 GCCCTTGCGGCCTTGGCCCCCTTACAGACTATGTGGT 696 
RESULT 40	
US-09-902-540-1194	
; Sequence 1194, Application US/09902540	
; Patent No. 6833447	
; GENERAL INFORMATION:	
; APPLICANT: Goldman, Barry S.	
; APPLICANT: Hinkle, Gregory J.	
; APPLICANT: Slater, Steven C.	
; APPLICANT: Wiegand, Roger C.	
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof	
; FILE REFERENCE: 38-10(15849)B	
; CURRENT APPLICATION NUMBER: US/09/902,540	
; CURRENT FILING DATE: 2001-07-10	
; PRIOR APPLICATION NUMBER: 60/217,883	
; PRIOR FILING DATE: 2000-07-10	
; NUMBER OF SEQ ID NOS: 16825	
; SEQ ID NO 1194	
; LENGTH: 21295	
; TYPE: DNA	
; ORGANISM: Myxococcus xanthus	
US-09-902-540-1194	
Query Match	48.6%; Score 20.4; DB 3; Length 21295;
Best Local Similarity	71.1%; Pred. No. 89;
Matches	27; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
QY	1 GCCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGT 38 
Db	15744 GCCCGAGCCTGGCTGGAGCGCTTTCAACGCCTATGACGT 15781 
RESULT 41	
US-08-086-428B-49	
; Sequence 49, Application US/08086428B	



ZIP: 10154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/468,570  
FILING DATE: 6-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/086,428  
FILING DATE: 29-JUN-1993  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: RICHARD W. BORK  
REGISTRATION NUMBER: 36,459  
REFERENCE/DOCKET NUMBER: 2026-4070US1  
TELEPHONE: (212) 758-4800  
TELEFAX: (212) 751-6849  
TELEX: 421792  
INFORMATION FOR SEQ ID NO: 49:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 576 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
ORGANISM: homosapiens  
INDIVIDUAL ISOLATE: SA7  
US-08-468-570-49

Query Match 47.6%; Score 20; DB 2; Length 576;  
Best Local Similarity 72.2%; Pred. No. 50;  
Matches 26; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
QY 2 CCCACGCTTGGCGGAGCCCTTACAAAGACTATGAAG 37  
Db 38 CCAATGATTCGCCGAACTCTTCATAGTCTATGAGG 73

RESULT 43  
US-08-290-665A-49  
Sequence 49, Application US/08290665A  
Patent No. 582852  
GENERAL INFORMATION:  
APPLICANT: BURKH, J., MILLER, R.H. AND  
PURCELL, R.H.  
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED  
AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND  
CORE GENES OF ISOLATES OF HEPATITIS C VIRUS  
TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS  
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE  
SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES  
NUMBER OF SEQUENCES: 263  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORGAN & FINNEGAN  
STREET: 345 PARK AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/290,665A  
FILING DATE: 15-AUG-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: RICHARD W. BORK

REGISTRATION NUMBER: 36,459  
REFERENCE/DOCKET NUMBER: 2026-4116  
TELEPHONE: (212) 758-4800  
TELEFAX: (212) 751-6849  
TELEX: 421792  
INFORMATION FOR SEQ ID NO: 49:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 576 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
ORGANISM: homosapiens  
INDIVIDUAL ISOLATE: SA7  
US-08-290-665A-49

Query Match 47.6%; Score 20; DB 2; Length 576;  
Best Local Similarity 72.2%; Pred. No. 50;  
Matches 26; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 2 CCCACGCTTGGCGGAGCCCTTACAAAGACTATGAAG 37  
Db 38 CCAATGATTCGCCGAACTCTTCATAGTCTATGAGG 73

RESULT 44  
US-08-466-601A-49  
Sequence 49, Application US/08466601A  
Patent No. 6572864  
GENERAL INFORMATION:  
APPLICANT: BURKH, J., MILLER, R.H. AND  
PURCELL, R.H.  
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED  
AMINO ACID SEQUENCES OF THE ENVELOPE 1 GENE  
TITLE OF INVENTION: OF 51 ISOLATES OF HEPATITIS C AND THE USE  
OF REAGENTS DERIVED FROM THESE SEQUENCES IN  
DIAGNOSTIC METHODS AND VACCINES  
NUMBER OF SEQUENCES: 160  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORGAN & FINNEGAN  
STREET: 345 PARK AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/466,601A  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/086,428  
FILING DATE: 29-JUN-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: RICHARD W. BORK  
REGISTRATION NUMBER: 36,459  
REFERENCE/DOCKET NUMBER: 2026-4070US2  
TELEPHONE: (212) 758-4800  
TELEFAX: (212) 751-6849  
TELEX: 421792  
INFORMATION FOR SEQ ID NO: 49:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 576 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

```
; ORIGINAL SOURCE:
; ORGANISM: homosapiens
; INDIVIDUAL ISOLATE: SA7
US-08-466-601A-49

Query Match          47.6%; Score 20; DB 3; Length 576;
Best Local Similarity 72.2%; Pred. No. 50;
Matches 26; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 2 CCACGCTTGGCGAGCCCTTACAAAGACTATGAAG 37
Db 38 CCAATGATTGCCGGAACCTTCCATAGTCTATGAGG 73

RESULT 45
PCT-US95-10398-49
; Sequence 49, Application PC/TUS9510398
; GENERAL INFORMATION:
; APPLICANT: BURKH, J., MILLER, R.H. AND
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
; NUMBER OF SEQUENCES: 263
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10398
; FILING DATE: 15-AUG-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/086,428
; FILING DATE: 29 JUNE 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/290/665
; FILING DATE: 15 AUGUST 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: RICHARD W. BORK
; REGISTRATION NUMBER: 36,459
; REFERENCE/DOCKET NUMBER: 2026-4116
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 49:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 576 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; ORIGINAL SOURCE:
; ORGANISM: homosapiens
; INDIVIDUAL ISOLATE: SA7
PCT-US95-10398-49

Query Match          47.6%; Score 20; DB 7; Length 576;
Best Local Similarity 72.2%; Pred. No. 50;
Matches 26; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 2 CCACGCTTGGCGAGCCCTTACAAAGACTATGAAG 37
Db 38 CCAATGATTGCCGGAACCTTCCATAGTCTATGAGG 73

us-10-600-816-30.rni

ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: SA7
US-08-466-601A-49

Query Match          47.6%; Score 20; DB 3; Length 576;
Best Local Similarity 72.2%; Pred. No. 50;
Matches 26; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 2 CCACGCTTGGCGAGCCCTTACAAAGACTATGAAG 37
Db 38 CCAATGATTGCCGGAACCTTCCATAGTCTATGAGG 73

RESULT 46
US-09-949-016-15770
; Sequence 15770, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15770
; LENGTH: 211049
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-15770

Query Match          47.6%; Score 20; DB 3; Length 211049;
Best Local Similarity 82.1%; Pred. No. 2.5e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GCCACGCTTGGCGAGCCCTTACAAAG 28
Db 113886 GCCACGCTTGGCTGAACCTTTTACAAAG 113913

RESULT 47
US-09-533-559-647/c
; Sequence 647, Application US/09533559
; Patent No. 6902887
; GENERAL INFORMATION:
; APPLICANT: Randy M. Berka
; APPLICANT: Michael W. Rey
; APPLICANT: Jeffrey R. Shuster
; APPLICANT: Sakari Kauppinen
; APPLICANT: Ib Groth Clausen
; APPLICANT: Peter Bjarke Olsen
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 5849.200-US
; CURRENT APPLICATION NUMBER: US/09/533,559
; CURRENT FILING DATE: 2000-03-22
; EARLIER APPLICATION NUMBER: 09/273,623
; EARLIER FILING DATE: 1999-03-22
; NUMBER OF SEQ ID NOS: 7860
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 647
; LENGTH: 546
; TYPE: DNA
; ORGANISM: Fusarium venenatum
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(546)
; OTHER INFORMATION: n = A,T,C or G
US-09-533-559-647

Query Match          47.1%; Score 19.8; DB 3; Length 546;
Best Local Similarity 67.5%; Pred. No. 60;
Matches 27; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAA 40
Db 113886 GCCACGCTTGGCTGAACCTTTTACAAAG 113913
```

```
Db 491 GCCCTTGCTAGAGGCGTTGACAAAGTCCAGGAAGTCA 452

RESULT 48
US-09-949-016-148482/c
; Sequence 148482, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 148482
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-148482

Query Match 47.1%; Score 19.8; DB 3; Length 601;
Best Local Similarity 77.4%; Pred. No. 62;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 CCCACGCTTGCGCCGAGCCCTTACAAAGACTA 32
Db 524 CCCACCTGGCCCAAGCCCAAACTCAGACTA 494

RESULT 49
US-09-949-016-148483/c
; Sequence 148483, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 148483
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-148483

Query Match 47.1%; Score 19.8; DB 3; Length 601;
Best Local Similarity 77.4%; Pred. No. 62;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 CCCACGCTTGCGCCGAGCCCTTACAAAGACTA 32
Db 260 CCCACCTGGCCCAAGCCCAAACTCAGACTA 230

RESULT 50
US-09-717-432-1/c
```

```
; Sequence 1, Application US/09717432
; Patent No. 6291223
; GENERAL INFORMATION:
; APPLICANT: ZHU, YUAN
; APPLICANT: LI, XIAOTONG
; APPLICANT: CHRISTIE, GARY
; APPLICANT: POWELL, DAVID J.
; TITLE OF INVENTION: Mouse Aspartic Secretase-1 (mASPl)
; FILE REFERENCE: GP-70663
; CURRENT APPLICATION NUMBER: US/09/717,432
; CURRENT FILING DATE: 2000-11-21
; PRIOR APPLICATION NUMBER: 60/166,974
; PRIOR FILING DATE: 1999-11-23
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1545
; TYPE: DNA
; ORGANISM: MUS MUSCULUS
US-09-717-432-1

Query Match 47.1%; Score 19.8; DB 3; Length 1545;
Best Local Similarity 69.2%; Pred. No. 80;
Matches 27; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1 GCCCAGCTTGGCGGAGCCCTTACAAAGACTATGAAGTA 39
Db 1269 GCCCACCCTCTGTGAGCTCTGTCAAAGACCACGTAGAA 1231

RESULT 51
US-09-912-484-1/c
; Sequence 1, Application US/09912484
; Patent No. 6358725
; GENERAL INFORMATION:
; APPLICANT: Christie, Gary
; APPLICANT: Li, Xiaotong
; APPLICANT: Powell, David J.
; APPLICANT: Zhu, Yuan
; TITLE OF INVENTION: Mouse Aspartic Secretase-1 (mASPl)
; FILE REFERENCE: GP-70663-D1
; CURRENT APPLICATION NUMBER: US/09/912,484
; CURRENT FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 60/166,974
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: 09/717,432
; PRIOR FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1545
; TYPE: DNA
; ORGANISM: MUS MUSCULUS
US-09-912-484-1

Query Match 47.1%; Score 19.8; DB 3; Length 1545;
Best Local Similarity 69.2%; Pred. No. 80;
Matches 27; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1 GCCCAGCTTGGCGGAGCCCTTACAAAGACTATGAAGTA 39
Db 1269 GCCCACCCTCTGTGAGCTCTGTCAAAGACCACGTAGAA 1231

RESULT 52
US-09-949-016-15910/c
; Sequence 15910, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
```

; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 15910  
; LENGTH: 8961  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-15910

Query Match 47.1%; Score 19.8; DB 3; Length 8961;  
Best Local Similarity 77.4%; Pred. No. 1.3e+02;  
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 CCCACGCTTGGCGAGCCCTTACAAAGACTA 32  
||||| ||||| ||||| ||||| |||||  
Db 5129 CCCACCTGGCCCAAGCCCAACTCAGACTA 5099

RESULT 53  
US-09-949-016-15306/c  
; Sequence 15306, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 15306  
; LENGTH: 9382  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-15306

Query Match 47.1%; Score 19.8; DB 3; Length 9382;  
Best Local Similarity 77.4%; Pred. No. 1.3e+02;  
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 CCCACGCTTGGCGAGCCCTTACAAAGACTA 32  
||||| ||||| ||||| ||||| |||||  
Db 1556 CCCAGGCTGCCCTAGCCCTTGCAATAATA 1526

RESULT 54  
US-09-949-016-13313/c  
; Sequence 13313, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768

; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13313  
; LENGTH: 75212  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-13313

Query Match 47.1%; Score 19.8; DB 3; Length 75212;  
Best Local Similarity 69.2%; Pred. No. 2.3e+02;  
Matches 27; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 3 CCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAA 41  
||||| ||||| ||||| ||||| |||||  
Db 29029 CCATGCTGGCCGAACACTATAATTCTATTATTGAA 28991

RESULT 55  
US-09-949-016-13314/c  
; Sequence 13314, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13314  
; LENGTH: 75212  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-13314

Query Match 47.1%; Score 19.8; DB 3; Length 75212;  
Best Local Similarity 69.2%; Pred. No. 2.3e+02;  
Matches 27; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 3 CCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAA 41  
||||| ||||| ||||| ||||| |||||  
Db 29029 CCATGCTGGCCGAACACTATAATTCTATTATTGAA 28991

RESULT 56  
US-09-949-016-13315/c  
; Sequence 13315, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0





```

; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH POLYOMYELITIS WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14083
; LENGTH: 199471
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(199471)
; OTHER INFORMATION: n = A,T,C or G
; US-09-949-016-14083

```

```
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 146770
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-146770

Query Match          46.7%; Score 19.6; DB 3; Length 601;
Best Local Similarity 73.5%; Pred. No. 76;
Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 9 TTGCGCGAGCCCTTACAAAGACTATGAAGTAAAG 42
   ||||| ||| ||| ||| ||| ||| ||| |||
Db 369 TTGCGCGAGCTTTTCATAAACTTGAAATATAG 336

RESULT 69
US-09-205-815B-49/c
; Sequence 49, Application US/09205815B
; Patent No. 6492509
; GENERAL INFORMATION:
; APPLICANT: Calgene, LLC
; TITLE OF INVENTION: Fatty Acyl-CoA: Fatty Alcohol Acyltransferases
; FILE REFERENCE: Fatty Acyl-CoA: Fatty Alcohol Acyltran
; CURRENT APPLICATION NUMBER: US/09/205,815B
; CURRENT FILING DATE: 1998-12-04
; PRIOR APPLICATION NUMBER: 09/092,562
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/048,651
; PRIOR FILING DATE: 1997-06-05
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: IBM PC; Windows NT 4.0; Microsoft Word For Windows 7.0a
; SEQ ID NO 49
; LENGTH: 1020
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; LOCATION: (1)..(1020)
; NAME/KEY: Gene
; OTHER INFORMATION: sequence of AT_WS7
; US-09-205-815B-49

Query Match          46.7%; Score 19.6; DB 3; Length 1020;
Best Local Similarity 73.5%; Pred. No. 87;
Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 5 ACCTTGGCGGAGCCCTTACAAAGACTATGAAGT 38
   ||||| ||| ||| ||| ||| ||| |||
Db 623 ACGATTGCCGAGAGACTTGGCGAGACTATGAGT 590

RESULT 70
5486473-1
; Patent No. 5486473
; APPLICANT: FUKITA, HIROYUKI; YOSHIDA, IWAO; TAKAGI, MITSUO;
; MANABE, SADAOKI; FUKAI, KONOSUKE
; TITLE OF INVENTION: A DNA CODING FOR A FLAVIVIRUS ANTIGEN
; . . . . . NUMBER OF SEQUENCES: 16
```

```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/194,049
; FILING DATE: 09-FEB-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 809,255
; FILING DATE: 18-DEC-1991
; APPLICATION NUMBER: 279,685
; FILING DATE: 05-DEC-1988
; APPLICATION NUMBER: 932,419
; FILING DATE: 19-NOV-1986
; SEQ ID NO: 1;
; LENGTH: 1500
; 5486473-1

Query Match          46.7%; Score 19.6; DB 10; Length 1500;
Best Local Similarity 73.5%; Pred. No. 97;
Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 CAGCTTGGCGAGCCCTTACAAAGACTATGAAG 37
   ||||| ||| ||| ||| ||| ||| |||
Db 1203 CACGCTAGGCAAGGCCCTTTTCAACAACTTTGAAG 1236

RESULT 71
US-08-956-171E-316
; Sequence 316, Application US/08956171E
; Patent No. 6593114
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; Gil H. Choi
; Patrick S. Dillon
; Craig A. Rosen
; Steven C. Barash
; Michael R. Fannon
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5256
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/956,171E
; FILING DATE: 20-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/009,861
; FILING DATE: January 5, 1996
; APPLICATION NUMBER: 08/781,986
; FILING DATE: January 3, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mark J. Hyman
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB248P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (240) 314-1224
; TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 316:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9310 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 316:
; US-08-956-171E-316
```



Query Match 46.7%; Score 19.6; DB 3; Length 9310;  
Best Local Similarity 73.5%; Pred. No. 1.6e+02;  
Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 9 TTGCGCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 3885 TTGCGGGGGCCCCAACAAAGAGCTGACGAAAG 3918

## RESULT 72

US-08-781-986A-316  
; Sequence 316, Application US/08781986A  
; Patent No. 6737248  
; GENERAL INFORMATION:  
; APPLICANT: Charles Kunsch  
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
; NUMBER OF SEQUENCES: 5255  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/781,986A  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Benson, Bob  
; REGISTRATION NUMBER: 30,446  
; REFERENCE/DOCKET NUMBER: PB248PP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 316:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 9310 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
US-08-781-986A-316

Query Match 46.7%; Score 19.6; DB 3; Length 9310;  
Best Local Similarity 73.5%; Pred. No. 1.6e+02;  
Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 9 TTGCGCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 3885 TTGCGGGGGCCCCAACAAAGAGCTGACGAAAG 3918

## RESULT 73

US-09-452-638-52  
; Sequence 52, Application US/09452638  
; Patent No. 6696281  
; GENERAL INFORMATION:  
; APPLICANT: Chambers, Thomas J.  
; APPLICANT: Monath, Thomas P.  
; APPLICANT: Guirakhoo, Farshad  
; TITLE OF INVENTION: Chimeric Flavivirus Vaccines  
; FILE REFERENCE: 06132/033004  
; CURRENT APPLICATION NUMBER: US/09/452,638  
; CURRENT FILING DATE: 1999-12-01  
; PRIOR APPLICATION NUMBER: US 09/121,587

; PRIOR FILING DATE: 1998-07-23  
; PRIOR APPLICATION NUMBER: PCT/US98/03894  
; PRIOR FILING DATE: 1998-03-02  
; PRIOR APPLICATION NUMBER: US 09/007,664  
; PRIOR FILING DATE: 1998-01-15  
; PRIOR APPLICATION NUMBER: US 08/807,445  
; PRIOR FILING DATE: 1997-02-28  
; NUMBER OF SEQ ID NOS: 85  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 52  
; LENGTH: 10892  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: derived from Yellow Fever virus and Japanese  
; OTHER INFORMATION: Encephalitis virus  
; NAME/KEY: CDS  
; LOCATION: (119)...(10381)  
US-09-452-638-52

Query Match 46.7%; Score 19.6; DB 3; Length 10892;  
Best Local Similarity 73.5%; Pred. No. 1.7e+02;  
Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 4 CACGCTTGCGCGAGCCCTTACAAAGACTATGAAG 37  
||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 2185 CACGCTGGGCAAGGCCTTTTCAACAACCTTTGAAG 2218

## RESULT 74

US-09-121-587A-12  
; Sequence 12, Application US/09121587A  
; Patent No. 6962708  
; GENERAL INFORMATION:  
; APPLICANT: Chambers, Thomas J.  
; APPLICANT: Guirakhoo, Farshad  
; APPLICANT: Monath, Thomas P.  
; TITLE OF INVENTION: CHIMERIC FLAVIVIRUS VACCINES  
; FILE REFERENCE: 06132/033003  
; CURRENT APPLICATION NUMBER: US/09/121,587A  
; CURRENT FILING DATE: 1998-07-23  
; PRIOR APPLICATION NUMBER: US98/03894  
; PRIOR FILING DATE: 1998-03-02  
; PRIOR APPLICATION NUMBER: 09/007,664  
; PRIOR FILING DATE: 1998-01-15  
; PRIOR APPLICATION NUMBER: 08/807,445  
; PRIOR FILING DATE: 1997-02-28  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12  
; LENGTH: 10892  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: derived from Japanese Encephalitis virus and  
; OTHER INFORMATION: Yellow Fever virus  
; NAME/KEY: CDS  
; LOCATION: (119)...(10381)  
US-09-121-587A-12

Query Match 46.7%; Score 19.6; DB 3; Length 10892;  
Best Local Similarity 73.5%; Pred. No. 1.7e+02;  
Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 4 CACGCTTGCGCGAGCCCTTACAAAGACTATGAAG 37  
||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 2185 CACGCTGGGCAAGGCCTTTTCAACAACCTTTGAAG 2218

## RESULT 75

US-09-949-002-832  
; Sequence 832, Application US/09949002  
; Patent No. 6900016

```
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; AND USES THEREOF
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 832
; LENGTH: 36677
; TYPE: DNA
; ORGANISM: Human
US-09-949-002-832
```

```
Query Match          46.7%; Score 19.6; DB 3; Length 36677;
Best Local Similarity 73.5%; Pred. No. 2.4e+02;
Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
```

```
QY 9 TTGCGCGAGCCCTTACAAAGACTATGAAGTAAAG 42
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 18546 TGGCGAGAGCCCTTAAGGAACAATGAAGTAAAG 18579.
```

```
RESULT 76
US-09-949-016-15863
; Sequence 15863, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15863
; LENGTH: 96340
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15863
```

```
Query Match          46.7%; Score 19.6; DB 3; Length 96340;
Best Local Similarity 73.5%; Pred. No. 3.1e+02;
Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
```

```
QY 9 TTGCGCGAGCCCTTACAAAGACTATGAAGTAAAG 42
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 36672 TTGCGCCAGTCTTTTCATAAATCTTTGAATATAG 36705
```

```
RESULT 77
US-09-949-016-13750/c
; Sequence 13750, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
```

```
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13750
; LENGTH: 132456
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(132456)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13750
```

```
Query Match          46.7%; Score 19.6; DB 3; Length 132456;
Best Local Similarity 66.7%; Pred. No. 3.3e+02;
Matches 28; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
```

```
QY 1 GCCCAGCGCTTGCGCGAGCCCTTACAAAGACTATGAAGTAAAG 42
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 85965 GCCCGCGCTCCGCCCGCTCCATAAAATGGCTAGGCTGTCAAG 85924
```

```
RESULT 78
US-09-949-016-15436/c
; Sequence 15436, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15436
; LENGTH: 199945
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15436
```

```
Query Match          46.7%; Score 19.6; DB 3; Length 199945;
Best Local Similarity 66.7%; Pred. No. 3.7e+02;
Matches 28; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
```

```
QY 1 GCCCAGCGCTTGCGCGAGCCCTTACAAAGACTATGAAGTAAAG 42
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 134050 GCCACTGTGGGCCCGCCAGCCCTTAGAGAGGCCCTGAAGCCCGAG 134009
```

```
RESULT 79
US-09-949-016-14588
; Sequence 14588, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
```

```
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14588
; LENGTH: 302604
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(302604)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14588

Query Match          46.7%; Score 19.6; DB 3; Length 302604;
Best Local Similarity 73.5%; Pred. No. 4.1e+02;
Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 7 GCTTGCCGAGCCCTTACAAGACTATGAAGTAA 40
Db 1278 GCTTCTGAGTCTCTACAAGACTCCGGAGTAA 1311

RESULT 80
US-09-949-016-14589
; Sequence 14589, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14589
; LENGTH: 302604
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(302604)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14589

Query Match          46.7%; Score 19.6; DB 3; Length 302604;
Best Local Similarity 73.5%; Pred. No. 4.1e+02;
Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 7 GCTTGCCGAGCCCTTACAAGACTATGAAGTAA 40
Db 1278 GCTTCTGAGTCTCTACAAGACTCCGGAGTAA 1311

RESULT 81
US-09-949-016-17119
; Sequence 17119, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
```

```
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17119
; LENGTH: 308362
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(308362)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17119

Query Match          46.7%; Score 19.6; DB 3; Length 308362;
Best Local Similarity 73.5%; Pred. No. 4.1e+02;
Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 7 GCTTGCCGAGCCCTTACAAGACTATGAAGTAA 40
Db 1094 GCTTCTGAGTCTCTACAAGACTCCGGAGTAA 1127

RESULT 82
US-09-621-976-16747
; Sequence 16747, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 16747
; LENGTH: 461
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-16747

Query Match          46.2%; Score 19.4; DB 3; Length 461;
Best Local Similarity 63.4%; Pred. No. 86;
Matches 26; Conservative 2; Mismatches 13; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGCCGAGCCCTTACAAGACTATGAAGTAA 41
Db 367 GCCCATGAGAGGCTGTGCTTTACAAGGASTCTCARTGATA 407

RESULT 83
US-09-949-016-95958/c
; Sequence 95958, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; CURRENT APPLICATION NUMBER: 60/231,498
```



RESULT 88  
US-09-949-016-14543  
; Sequence 14543, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14543  
; LENGTH: 387902  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(387902)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-14543  
Query Match 46.2%; Score 19.4; DB 3; Length 387902;  
Best Local Similarity 70.3%; Pred. No. 5.3e+02;  
Matches 26; Conservative 0; Mismatches 11; Indels 0; Gaps 0;  
Qy 5 ACGTTGGCCGAGCCCTTACAAGACTATGAAGTAA 41  
Db 28249 ACTCTGGCCAGTAAAGACGCGACTATGAAAAA 28285  
RESULT 89  
US-09-949-016-12557  
; Sequence 12557, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12557  
; LENGTH: 421883  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(421883)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-12557  
Query Match 46.2%; Score 19.4; DB 3; Length 421883;  
Best Local Similarity 70.3%; Pred. No. 5.3e+02;  
Matches 26; Conservative 0; Mismatches 11; Indels 0; Gaps 0;  
Qy 5 ACGTTGGCCGAGCCCTTACAAGACTATGAAGTAA 41

Db 28249 ACTCTGGCCAGTAAAGACGCGACTATGAAAAA 28285  
RESULT 90  
US-09-621-976-16083  
; Sequence 16083, Application US/09621976  
; Patent No. 6639063  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Jobert, S.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
; FILE REFERENCE: GENSET.054PR2  
; CURRENT APPLICATION NUMBER: US/09/621,976  
; CURRENT FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 19335  
; SOFTWARE: Patent.pm  
; SEQ ID NO 16083  
; LENGTH: 215  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-621-976-16083  
Query Match 45.7%; Score 19.2; DB 3; Length 215;  
Best Local Similarity 75.0%; Pred. No. 86;  
Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
Qy 9 TTGCGCGAGCCCTTACAAGACTATGAAGTAA 40  
Db 118 TTGCGCAAAACCTTCAAGGATATGCAGAA 149  
RESULT 91  
US-09-252-991A-8407  
; Sequence 8407, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 8407  
; LENGTH: 450  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-8407  
Query Match 45.7%; Score 19.2; DB 3; Length 450;  
Best Local Similarity 75.0%; Pred. No. 1.1e+02;  
Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
Qy 3 CCACGGTTGGCCGAGCCCTTACAAGACTATG 34  
Db 107 CAAGCGCTGGCCGAGCCGCTCCATCGACTATG 138  
RESULT 92  
US-09-878-281A-153  
; Sequence 153, Application US/09878281A  
; Patent No. 6762024  
; GENERAL INFORMATION:  
; APPLICANT: Innogenetics N.V.  
; TITLE OF INVENTION: New sequences of hepatitis C virus genotypes for diagnosis, proph  
; TITLE OF INVENTION: and therapy  
; FILE REFERENCE: 35



```

; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 47026
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-47026

```

Query Match 45.7%; Score 19.2; DB 3; Length 601;  
Best Local Similarity 67.5%; Pred. No. 1.1e+02;  
Matches 27; Conservative 0; Mismatches 13; Indels

Qy 3 CCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
| | | | | | | | | | | | | | | | | | | | | |  
Db 50 CAACTCTTGGGTACCTGTACAGAGACTAATGTGAAG 89

RESULT 98  
US-08-612-973-31  
; Sequence 31, Application US/08612973  
; Patent No. 6150134  
; GENERAL INFORMATION:  
; APPLICANT: MAERTENS, GEERT  
; APPLICANT: BOSMAN, FONS  
; APPLICANT: DE MARTYNOFF, GUY  
; APPLICANT: BUYSSE, MARIE-ANGE  
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE  
; TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE  
; NUMBER OF SEQUENCES: 111  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NIXON & VANDERHYE P. C.  
;

Query Match	45.7%;	Score 19.2;	DB 3;	Length 630;
Best Local Similarity	67.5%;	Pred. No. 1.2e+02;		

	Matches	27;	Conservative	0;	Mismatches	13;	Indels	0;	Gaps	0;
Qy	2	CCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAA	41							
Db	257	CCAAATGATGCCCAAACTCTTCCATAGTCTATGAGGCAGA	296							

```

RESULT 99
US-08-927-597-31
; Sequence 31, Application US/08927597
; Patent No. 6245503
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: BOSMAN, FONS
; APPLICANT: DE MARTYNOFF, GUY
; APPLICANT: BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P. C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/927,597
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 630 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..627
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 1..624
; US-08-927-597-31

```

Query Match	45.7%	Score 19.2;	DB 3;	Length 630;
Best Local Similarity	67.5%;	Pred. No. 1.2e+02;		
Matches	27; Conservative	0; Mismatches	13; Indels	0; Gaps 0;
Qy . . .	2	CCACGCTTGGCGAGCCTTTACAAGACTATGAAGTAA	41	
Dd	257	CCAATGATTGCCAATACTTTTCATAGTGCTATCAGGCAGA	296	

RESULT 100  
US-08-928-757-31

```
; Sequence 31, Application US/08928757
; Patent No. 6890737
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
;             BOSMAN, FONS
;             DE MARTYNOFF, GUY
;             BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
;                   PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA: US/08/928,757
; FILING DATE: 12-Sep-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 630 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..627
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 1..624
; SEQUENCE DESCRIPTION: SEQ ID NO: 31:
US-08-928-757-31
```

```
Query Match          45.7%; Score 19.2; DB 3; Length 630;
Best Local Similarity 67.5%; Pred.No.1.2e+02;
Matches 27; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
```

```
QY      2  CCCACGCTGGCGGAGCCCTTACAAAGACTATGAAGTAA 41
Db      257  CCAATGATTGCCCAAACTCTTCATAGTCTATGAGGAGA 296
```

```
Search completed: June 10, 2006, 15:37:12
Job time : 150.15 secs
```



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2006, 15:26:46 ; Search time 1608.6 Seconds  
(without alignments)  
320.826 Million cell updates/sec

Title: US-10-600-816-30

Perfect score: 42

Sequence: 1 gccacgctggccgagccc.....acaaagactgaagtaag 42

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 18892170 seqs, 6143817638 residues

Total number of hits satisfying chosen parameters: 37784340

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

Database : Published Applications NA Main:\*

- 1: /EMC\_Celerra\_SID33/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /EMC\_Celerra\_SID33/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 3: /EMC\_Celerra\_SID33/ptodata/2/pubpna/US09A\_PUBCOMB.seq:\*
- 4: /EMC\_Celerra\_SID33/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*
- 5: /EMC\_Celerra\_SID33/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*
- 6: /EMC\_Celerra\_SID33/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*
- 7: /EMC\_Celerra\_SID33/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*
- 8: /EMC\_Celerra\_SID33/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*
- 9: /EMC\_Celerra\_SID33/ptodata/2/pubpna/US10D\_PUBCOMB.seq:\*
- 10: /EMC\_Celerra\_SID33/ptodata/2/pubpna/US10E\_PUBCOMB.seq:\*
- 11: /EMC\_Celerra\_SID33/ptodata/2/pubpna/US10F\_PUBCOMB.seq:\*
- 12: /EMC\_Celerra\_SID33/ptodata/2/pubpna/US10G\_PUBCOMB.seq:\*
- 13: /EMC\_Celerra\_SID33/ptodata/2/pubpna/US11A\_PUBCOMB.seq:\*
- 14: /EMC\_Celerra\_SID33/ptodata/2/pubpna/US11B\_PUBCOMB.seq:\*
- 15: /EMC\_Celerra\_SID33/ptodata/2/pubpna/US11C\_PUBCOMB.seq:\*
- 16: /EMC\_Celerra\_SID33/ptodata/2/pubpna/US11D\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42	100.0	42	8	US-10-600-816-30
2	42	100.0	497	6	US-10-066-543-1937
3	42	100.0	552	6	US-10-066-543-1811
4	42	100.0	620	3	US-09-969-034-2222
5	42	100.0	634	3	US-09-969-034-3393
6	42	100.0	642	7	US-10-125-968-701
7	42	100.0	1071	3	US-10-712-615-134
8	42	100.0	1212	3	US-09-866-050A-249
9	42	100.0	1212	6	US-10-152-661-249
10	42	100.0	1228	7	US-10-313-542-223
11	42	100.0	1400	13	US-11-060-756-2418
12	42	100.0	1400	13	US-11-060-756-6690
13	42	100.0	1460	10	US-10-936-626-64
14	42	100.0	1460	10	US-10-938-061-64
15	42	100.0	1619	7	US-10-224-289-5
16	42	100.0	1619	9	US-10-935-190-43
17	42	100.0	2297	9	US-10-775-920-13
					Sequence 30, Appl
					Sequence 1937, Ap
					Sequence 1811, Ap
					Sequence 2222, Ap
					Sequence 3393, Ap
					Sequence 701, App
					Sequence 134, App
					Sequence 249, App
					Sequence 223, App
					Sequence 2418, Ap
					Sequence 6690, Ap
					Sequence 64, Appl
					Sequence 5, Appli
					Sequence 43, Appl
					Sequence 13, Appl

7	US-10-224-289-3	2302	100.0	42	18
8	US-10-240-425-405	2302	100.0	42	19
9	US-10-775-920-9	2302	100.0	42	20
10	US-10-510-507-2	2302	100.0	42	21
9	US-10-775-920-12	2305	100.0	42	22
6	US-10-176-847-59	2316	100.0	42	23
15	US-11-080-991-59	2316	100.0	42	24
6	US-10-775-920-11	2446	100.0	42	25
6	US-10-225-567A-453	2456	100.0	42	26
7	US-10-269-909-63	2456	100.0	42	27
7	US-10-269-909-64	2456	100.0	42	28
7	US-10-295-027-619	2456	100.0	42	29
8	US-10-600-816-2	2456	100.0	42	30
8	US-10-600-816-18	2456	100.0	42	31
9	US-10-775-920-10	2456	100.0	42	32
10	US-10-936-626-40	2456	100.0	42	33
10	US-10-938-061-40	2456	100.0	42	34
16	US-11-169-041-32	2456	100.0	42	35
7	US-10-264-049-834	2593	100.0	42	36
6	US-10-198-846-10424	4239	100.0	42	37
10	US-10-505-486-196	1788	96.2	40.4	38
8	US-10-600-816-33	71.4	30	30	39
13	US-11-060-756-43361	59.5	25	25	40
9	US-10-459-155-7	480	54.3	22.8	41
9	US-10-459-155-12	525	54.3	22.8	42
9	US-10-459-155-9	1500	54.3	22.8	43
9	US-10-459-155-21	1545	54.3	22.8	44
4	US-09-925-065A-691738	1540	53.3	22.4	45
4	US-09-925-065A-691739	1540	53.3	22.4	46
4	US-09-925-065A-691740	1540	53.3	22.4	47
5	US-09-925-065A-691738	1540	53.3	22.4	48
5	US-09-925-065A-691739	1540	53.3	22.4	49
5	US-09-925-065A-691740	1540	53.3	22.4	50
3	US-09-917-800A-76	584	52.9	22.2	51
4	US-09-925-065A-49510	2548	52.9	22.2	52
5	US-09-925-065A-49510	2548	52.9	22.2	53
12	US-10-301-480-150748	2548	52.9	22.2	54
12	US-10-301-480-764157	2548	52.9	22.2	55
7	US-10-017-161-455	1339	51.9	21.8	56
7	US-10-292-798-409	1339	51.9	21.8	57
6	US-10-044-090-126	4826	51.9	21.8	58
6	US-10-084-817-304	4826	51.9	21.8	59
3	US-09-764-847-10135	17849	51.9	21.8	60
3	US-09-764-891-13119	17849	51.9	21.8	61
6	US-10-092-154-1315	17849	51.9	21.8	62
3	US-09-764-847-1313	17862	51.9	21.8	63
3	US-09-764-891-10138	17862	51.9	21.8	64
6	US-10-092-154-1313	17862	51.9	21.8	65
13	US-11-097-143-22855	21748	51.9	21.8	66
9	US-10-476-991-3	70000	51.4	21.6	67
13	US-11-097-143-1964	2356	51.0	21.4	68
16	US-11-096-568A-26327	1227	50.5	21.2	69
10	US-10-450-763-28218	1446	50.5	21.2	70
7	US-10-369-493-36568	1555	50.5	21.2	71
9	US-10-425-115-130787	3141	50.5	21.2	72
3	US-09-814-353-3621	527	50.0	21	73
3	US-09-814-353-9935	527	50.0	21	74
3	US-09-814-353-5769	559	50.0	21	75
3	US-09-814-353-12052	559	50.0	21	76
3	US-09-814-353-4193	579	50.0	21	77
3	US-09-814-353-10499	579	50.0	21	78
3	US-09-814-353-18436	625	50.0	21	79
3	US-09-814-353-18787	632	50.0	21	80
3	US-09-814-353-20633	1096	50.0	21	81
3	US-09-814-353-20810	1386	50.0	21	82
3	US-09-814-353-20790	2145	50.0	21	83
7	US-10-006-285-420	4605	50.0	21	84
8	US-10-723-860-5576	96589	50.0	21	85
8	US-10-052-482-214	560	49.5	20.8	86
4	US-09-925-065A-148760	560	49.5	20.8	87
5	US-09-925-065A-148761	560	49.5	20.8	88
5	US-09-925-065A-148760	560	49.5	20.8	89
5	US-09-925-065A-148761	560	49.5	20.8	90















```

c 967 18.6 44.3 4806 13 US-11-097-143-11887
968 18.6 44.3 5082 13 US-11-097-143-4180
969 18.6 44.3 6975 13 US-11-097-143-11656
970 18.6 44.3 10343 13 US-11-097-143-18610
c 971 18.6 44.3 24678 9 US-10-741-600-17702
c 972 18.6 44.3 50000 6 US-10-152-724A-22
973 18.6 44.3 96583 15 US-11-117-187-203
c 974 18.6 44.3 96595 7 US-10-034-650-34
975 18.6 44.3 117329 10 US-10-987-384-1
c 976 18.6 44.3 127369 6 US-10-087-192-238
977 18.6 44.3 180654 15 US-11-121-086-58
978 18.6 44.3 197241 16 US-11-114-798-47
979 18.6 44.3 215221 6 US-10-087-192-1360
980 18.6 44.3 218336 9 US-10-473-939-10
c 981 18.6 44.3 310122 9 US-10-417-375-1
982 18.6 44.3 382256 10 US-10-820-226-1
983 18.6 44.3 382259 13 US-11-029-984-1
c 984 18.6 44.3 559686 5 US-09-824-575A-3
c 985 18.6 44.3 653458 10 US-10-461-862-4
986 18.6 44.3 1223197 6 US-10-027-632-179264
987 18.6 44.3 1223197 7 US-10-027-632-179264
988 18.6 44.3 201 10 US-10-995-561-64681
c 989 18.4 43.8 300 10 US-10-779-543-3237
c 990 18.4 43.8 369 16 US-11-102-512-110
991 18.4 43.8 371 9 US-10-425-115-82970
992 18.4 43.8 444 12 US-10-301-480-9374
993 18.4 43.8 444 12 US-10-301-480-9375
994 18.4 43.8 444 12 US-10-301-480-622783
995 18.4 43.8 444 12 US-10-301-480-622784
996 18.4 43.8 445 4 US-09-925-065A-248267
997 18.4 43.8 445 5 US-09-925-065A-248267
c 998 18.4 43.8 448 3 US-09-960-352-8299
999 18.4 43.8 451 12 US-10-301-480-328272
1000 18.4 43.8 451 12 US-10-301-480-941681

```

# ALIGNMENTS

```

RESULT 1
US-10-600-816-30
; Sequence 30, Application US/10600816
; Publication No. US20040121362A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: IDENTIFICATION AND MODULATION OF A G-PROTEIN COUPLED RECEPTOR
; FILE REFERENCE: D0251 NP
; CURRENT APPLICATION NUMBER: US/10/600,816
; PRIOR FILING DATE: 2003-06-20
; PRIOR APPLICATION NUMBER: U.S. 60/390,850
; PRIOR FILING DATE: 2002-06-20
; PRIOR APPLICATION NUMBER: U.S. 60/407,006
; PRIOR FILING DATE: 2002-08-29
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 30
; LENGTH: 42
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-600-816-30

```

```

Query Match 100.0%; Score 42; DB 8; Length 42;
Best Local Similarity 100.0%; Pred. No. 4.4e-08;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 42
|||||
Db 1 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 42
|||||

```

RESULT 2

```

US-10-066-543-1937
; Sequence 1937, Application US/10066543
; Publication No. US20030087818A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Indrias, Carol Yoseph
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Carter, Darrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Smith, Carole L.
; APPLICANT: Durham, Margarita
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.563
; CURRENT APPLICATION NUMBER: US/10/066,543
; CURRENT FILING DATE: 2002-01-31
; NUMBER OF SEQ ID NOS: 3417
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1937
; LENGTH: 497
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-066-543-1937

```

```

Query Match 100.0%; Score 42; DB 6; Length 497;
Best Local Similarity 100.0%; Pred. No. 7.8e-08;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 42
|||||
Db 344 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 385
|||||

```

RESULT 3

```

US-10-066-543-1811/c
; Sequence 1811, Application US/10066543
; Publication No. US20030087818A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Indrias, Carol Yoseph
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Carter, Darrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Smith, Carole L.
; APPLICANT: Durham, Margarita
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.563
; CURRENT APPLICATION NUMBER: US/10/066,543
; CURRENT FILING DATE: 2002-01-31
; NUMBER OF SEQ ID NOS: 3417
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1811
; LENGTH: 552
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-066-543-1811

```

```

Query Match 100.0%; Score 42; DB 6; Length 552;
Best Local Similarity 100.0%; Pred. No. 8e-08;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 42
|||||
Db 154 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 113
|||||

```



## RESULT 4

US-09-969-034-2222/c  
; Sequence 2222, Application US/09969034  
; Publication No. US20040110668A1  
; GENERAL INFORMATION:  
; APPLICANT: Burgess, Christopher C.  
; APPLICANT: Astle, Jon H.  
; APPLICANT: Carroll, Eddie III  
; APPLICANT: Catino, Theodore J.  
; APPLICANT: Dwivedi, Poorima  
; APPLICANT: Molino, Gary A.  
; APPLICANT: Thiagalingam, Arunthathi  
; APPLICANT: Lewis, Marcia E.  
; TITLE OF INVENTION: Nucleic Acid Sequences Differentially  
; FILE REFERENCE: 1657/1032  
; CURRENT APPLICATION NUMBER: US/09/969,034  
; CURRENT FILING DATE: 2001-10-02  
; PRIOR APPLICATION NUMBER: 60/237,271  
; PRIOR FILING DATE: 2000-02-10  
; NUMBER OF SEQ ID NOS: 4494  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2222  
; LENGTH: 620  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 491, 500, 517, 530, 545, 547, 553, 558, 570, 599, 620  
; OTHER INFORMATION: n = A,T,C or G  
US-09-969-034-2222

Query Match 100.0%; Score 42; DB 3; Length 620;  
Best Local Similarity 100.0%; Pred. No. 8.2e-08;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGGCGAGCCCTTACAAGACTATGAAGTAAAG 42  
Db 154 GCCACGCTTGGCGAGCCCTTACAAGACTATGAAGTAAAG 113

## RESULT 5

US-09-969-034-3393/c  
; Sequence 3393, Application US/09969034  
; Publication No. US20040110668A1  
; GENERAL INFORMATION:  
; APPLICANT: Burgess, Christopher C.  
; APPLICANT: Astle, Jon H.  
; APPLICANT: Carroll, Eddie III  
; APPLICANT: Catino, Theodore J.  
; APPLICANT: Dwivedi, Poorima  
; APPLICANT: Molino, Gary A.  
; APPLICANT: Thiagalingam, Arunthathi  
; APPLICANT: Lewis, Marcia E.  
; TITLE OF INVENTION: Nucleic Acid Sequences Differentially  
; FILE REFERENCE: 1657/1032  
; CURRENT APPLICATION NUMBER: US/09/969,034  
; CURRENT FILING DATE: 2001-10-02  
; PRIOR APPLICATION NUMBER: 60/237,271  
; PRIOR FILING DATE: 2000-02-10  
; NUMBER OF SEQ ID NOS: 4494  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3393  
; LENGTH: 634  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 363, 470, 472, 516, 523, 531, 551, 567, 573, 588, 594, 613,  
; LOCATION: 623

; OTHER INFORMATION: n = A,T,C or G  
US-09-969-034-3393

Query Match 100.0%; Score 42; DB 3; Length 634;  
Best Local Similarity 100.0%; Pred. No. 8.3e-08;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GCCACGCTTGGCGAGCCCTTACAAGACTATGAAGTAAAG 42  
Db 158 GCCACGCTTGGCGAGCCCTTACAAGACTATGAAGTAAAG 117

## RESULT 6

US-10-125-968-701/c  
; Sequence 701, Application US/10125968  
; Publication No. US20030215805A1  
; GENERAL INFORMATION:  
; APPLICANT: Lillie, James  
; APPLICANT: Palermo, Adam  
; APPLICANT: Wang, Youzhen  
; APPLICANT: Steinmann, Kathleen  
; APPLICANT: Elias, Josh  
; APPLICANT: Mertens, Maureen  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND  
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
; FILE REFERENCE: MRI-032  
; CURRENT APPLICATION NUMBER: US/10/125,968  
; CURRENT FILING DATE: 2002-04-19  
; PRIOR APPLICATION NUMBER: US 60/285,163  
; PRIOR FILING DATE: 2001-04-20  
; NUMBER OF SEQ ID NOS: 1417  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 701  
; LENGTH: 642  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 72, 76, 77, 99, 101, 423, 445, 468, 489, 557, 566, 633  
; OTHER INFORMATION: n = A,T,C or G  
US-10-125-968-701

Query Match 100.0%; Score 42; DB 7; Length 642;  
Best Local Similarity 100.0%; Pred. No. 8.3e-08;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGGCGAGCCCTTACAAGACTATGAAGTAAAG 42  
Db 184 GCCACGCTTGGCGAGCCCTTACAAGACTATGAAGTAAAG 143

## RESULT 7

US-10-712-615-134  
; Sequence 134, Application US/10712615  
; Publication No. US20040214317A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: A NOVEL HUMAN G-PROTEIN COUPLED RECEPTOR, HGPRBMV8, EXPRESSED  
; FILE REFERENCE: D0047A-CIP  
; CURRENT APPLICATION NUMBER: US/10/712,615  
; CURRENT FILING DATE: 2003-11-13  
; PRIOR APPLICATION NUMBER: U.S. 09/992,238  
; PRIOR FILING DATE: 2001-11-14  
; PRIOR APPLICATION NUMBER: U.S. 60/248,285  
; PRIOR FILING DATE: 2000-11-14  
; PRIOR APPLICATION NUMBER: U.S. 60/268,581  
; PRIOR FILING DATE: 2001-02-14  
; PRIOR APPLICATION NUMBER: U.S. 60/308,285  
; PRIOR FILING DATE: 2001-07-27  
; PRIOR APPLICATION NUMBER: U.S. 60/317,166  
; PRIOR FILING DATE: 2001-09-04

; NUMBER OF SEQ ID NOS: 134  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 134  
; LENGTH: 1071  
; TYPE: DNA  
; ORGANISM: dna;Homo sapiens  
US-10-712-615-134

Query Match 100.0%; Score 42; DB 9; Length 1071;  
Best Local Similarity 100.0%; Pred. No. 9.4e-08;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACGCTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
|||||  
Db 1018 GCCACGCTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1059

## RESULT 8

US-09-866-050A-249  
; Sequence 249, Application US/09866050A  
; Publication No. US20030040471A1  
; GENERAL INFORMATION:  
; APPLICANT: Watson, James D.  
; APPLICANT: Strachan, Lorna  
; APPLICANT: Sleeman, Matthew  
; APPLICANT: Onrust, Rene  
; APPLICANT: Murison, James G.  
; APPLICANT: Kumble, Krishanand D.  
; TITLE OF INVENTION: Compositions Isolated From Skin Cells  
; FILE REFERENCE: 11000.1011C4U  
; CURRENT APPLICATION NUMBER: US/09/866,050A  
; CURRENT FILING DATE: 2001-05-24  
; NUMBER OF SEQ ID NOS: 725  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 249  
; LENGTH: 1212  
; TYPE: DNA  
; ORGANISM: Human  
US-09-866-050A-249

Query Match 100.0%; Score 42; DB 3; Length 1212;  
Best Local Similarity 100.0%; Pred. No. 9.6e-08;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACGCTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
|||||  
Db 990 GCCACGCTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1031

## RESULT 9

US-10-152-661-249  
; Sequence 249, Application US/10152661  
; Publication No. US20030022835A1  
; GENERAL INFORMATION:  
; APPLICANT: Watson, James D.  
; APPLICANT: Strachan, Lorna  
; APPLICANT: Sleeman, Matthew  
; APPLICANT: Onrust, Rene  
; APPLICANT: Murison, James G.  
; APPLICANT: Kumble, Krishanand D.  
; TITLE OF INVENTION: Compositions Isolated From Skin Cells  
; FILE REFERENCE: 11000.1011C5  
; CURRENT APPLICATION NUMBER: US/10/152,661  
; CURRENT FILING DATE: 2002-05-20  
; PRIOR APPLICATION NUMBER: 09/866,050  
; PRIOR FILING DATE: 2001-05-24  
; PRIOR APPLICATION NUMBER: 60/221,232  
; PRIOR FILING DATE: 2000-07-25  
; PRIOR APPLICATION NUMBER: 60/206,650  
; PRIOR FILING DATE: 2000-05-24  
; PRIOR APPLICATION NUMBER: 09/312,283

; PRIOR FILING DATE: 1999-05-14  
; PRIOR APPLICATION NUMBER: PCT/NZ99/00051  
; PRIOR FILING DATE: 1999-04-29  
; PRIOR APPLICATION NUMBER: 09/188,930  
; PRIOR FILING DATE: 1998-11-09  
; PRIOR APPLICATION NUMBER: 09/069,726  
; PRIOR FILING DATE: 1998-04-29  
; NUMBER OF SEQ ID NOS: 725  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 249  
; LENGTH: 1212  
; TYPE: DNA  
; ORGANISM: Human  
US-10-152-661-249

Query Match 100.0%; Score 42; DB 6; Length 1212;  
Best Local Similarity 100.0%; Pred. No. 9.6e-08;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACGCTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
|||||  
Db 990 GCCACGCTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1031

## RESULT 10

US-10-313-542-223  
; Sequence 223, Application US/10313542  
; Publication No. US20030120057A1  
; GENERAL INFORMATION:  
; APPLICANT: Roopa, Reddy  
; APPLICANT: Guegler, Karl, J.  
; APPLICANT: Au-Young, Janice  
; TITLE OF INVENTION: COMPOSITION FOR DETECTION OF GENES ENCODING MEMBRANE-ASSOCIATED P  
; FILE REFERENCE: PA-0013 US  
; CURRENT APPLICATION NUMBER: US/10/313,542  
; CURRENT FILING DATE: 2002-12-05  
; PRIOR APPLICATION NUMBER: US/09/495,050  
; PRIOR FILING DATE: 2000-01-31  
; PRIOR APPLICATION NUMBER: 60/118,318  
; PRIOR FILING DATE: 1999-02-01  
; NUMBER OF SEQ ID NOS: 305  
; SOFTWARE: PERL Program  
; SEQ ID NO 223  
; LENGTH: 1228  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. US20030120057A1 2681738CT1  
US-10-313-542-223

Query Match 100.0%; Score 42; DB 7; Length 1228;  
Best Local Similarity 100.0%; Pred. No. 9.7e-08;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACGCTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
|||||  
Db 541 GCCACGCTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 582

## RESULT 11

US-11-060-756-2418  
; Sequence 2418, Application US/11060756  
; Publication No. US20050221354A1  
; GENERAL INFORMATION:  
; APPLICANT: Wyeth  
; APPLICANT: Mounts, William Martin  
; TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug  
; FILE REFERENCE: AM101083 (031896-042000)  
; CURRENT APPLICATION NUMBER: US/11/060,756  
; CURRENT FILING DATE: 2005-02-18  
; NUMBER OF SEQ ID NOS: 303284

```
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2418
; LENGTH: 1400
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-11-060-756-2418

Query Match      100.0%; Score 42; DB 13; Length 1400;
Best Local Similarity 100.0%; Pred. No. 1e-07;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 42
    |||||
Db 215 GCCCAGCGTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 256

RESULT 12
US-11-060-756-6690
; Sequence 6690, Application US/11060756
; Publication No. US20050221354A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William Martin
; TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
; FILE REFERENCE: AM101083 (031896-042000)
; CURRENT APPLICATION NUMBER: US/11/060,756
; CURRENT FILING DATE: 2005-02-18
; NUMBER OF SEQ ID NOS: 303284
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6690
; LENGTH: 1400
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-11-060-756-6690

Query Match      100.0%; Score 42; DB 13; Length 1400;
Best Local Similarity 100.0%; Pred. No. 1e-07;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 42
    |||||
Db 215 GCCCAGCGTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 256

RESULT 13
US-10-936-626-64
; Sequence 64, Application US/10936626
; Publication No. US20050106644A1
; GENERAL INFORMATION:
; APPLICANT: Cairns, Belinda
; APPLICANT: Chen, Ruihuan
; APPLICANT: Frantz, Gretchen
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Koepfen, Hartmut
; APPLICANT: Phillips, Heidi S.
; APPLICANT: Polakis, Paul
; APPLICANT: Spencer, Susan D.
; APPLICANT: Smith, Victoria
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wu, Thomas D.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: Compositions and Methods for the Diagnosis and
; FILE REFERENCE: P5001R1P1
; CURRENT APPLICATION NUMBER: US/10/936,626
; CURRENT FILING DATE: 2004-09-08
; PRIOR APPLICATION NUMBER: US 10/872,991
; PRIOR FILING DATE: 2004-06-21
; PRIOR APPLICATION NUMBER: US 10/872,972
; PRIOR FILING DATE: 2004-06-21
; PRIOR APPLICATION NUMBER: US 10/241,220
; PRIOR FILING DATE: 2002-09-11
; PRIOR APPLICATION NUMBER: US 10/177,488
; PRIOR FILING DATE: 2002-06-19
; PRIOR APPLICATION NUMBER: US 60/299,500
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 60/301,880
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/323,268
; PRIOR FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US 60/557,116
; PRIOR FILING DATE: 2004-03-26
; PRIOR APPLICATION NUMBER: US 60/598,899
; PRIOR FILING DATE: 2004-06-21
; PRIOR APPLICATION NUMBER: US 10/241,220
; PRIOR FILING DATE: 2002-09-11
; SEQ ID NO 64
```

```
; PRIOR APPLICATION NUMBER: US 10/177,488
; PRIOR FILING DATE: 2002-06-19
; PRIOR APPLICATION NUMBER: US 60/299,500
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 60/301,880
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/323,268
; PRIOR FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US 60/557,116
; PRIOR FILING DATE: 2004-03-26
; PRIOR APPLICATION NUMBER: US 60/598,899
; NUMBER OF SEQ ID NOS: 154
; SEQ ID NO 64
; LENGTH: 1460
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-936-626-64

Query Match      100.0%; Score 42; DB 10; Length 1460;
Best Local Similarity 100.0%; Pred. No. 1e-07;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 42
    |||||
Db 1145 GCCCAGCGTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 1186

RESULT 14
US-10-938-061-64
; Sequence 64, Application US/10938061
; Publication No. US20050107595A1
; GENERAL INFORMATION:
; APPLICANT: Cairns, Belinda
; APPLICANT: Chen, Ruihuan
; APPLICANT: Frantz, Gretchen
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Koepfen, Hartmut
; APPLICANT: Phillips, Heidi S.
; APPLICANT: Polakis, Paul
; APPLICANT: Spencer, Susan D.
; APPLICANT: Smith, Victoria
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wu, Thomas D.
; APPLICANT: Zhang, Zemin
; APPLICANT: Sakanaka, Chie
; APPLICANT: Chuntharapal, Anan
; APPLICANT: Reed Chae J.
; TITLE OF INVENTION: Compositions and Methods for the Diagnosis and
; FILE REFERENCE: P5001R1P1B
; CURRENT APPLICATION NUMBER: US/10/938,061
; CURRENT FILING DATE: 2004-09-10
; PRIOR APPLICATION NUMBER: US 10/872,991
; PRIOR FILING DATE: 2004-06-21
; PRIOR APPLICATION NUMBER: US 10/872,972
; PRIOR FILING DATE: 2004-06-21
; PRIOR APPLICATION NUMBER: US 10/241,220
; PRIOR FILING DATE: 2002-09-11
; PRIOR APPLICATION NUMBER: US 10/177,488
; PRIOR FILING DATE: 2002-06-19
; PRIOR APPLICATION NUMBER: US 60/299,500
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 60/301,880
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/323,268
; PRIOR FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US 60/557,116
; PRIOR FILING DATE: 2004-03-26
; PRIOR APPLICATION NUMBER: US 60/598,899
; NUMBER OF SEQ ID NOS: 154
; SEQ ID NO 64
```

```
; LENGTH: 1460
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-938-061-64

Query Match      100.0%; Score 42; DB 10; Length 1460;
Best Local Similarity 100.0%; Pred. No. 1e-07;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42
   |||||||
Db 1145 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1186

RESULT 15
US-10-224-289-5
; Sequence 5, Application US/10224289
; Publication No. US20030207288A1
; GENERAL INFORMATION:
; APPLICANT: LEWIN, DAVID A.
; TITLE OF INVENTION: GPCR-LIKE RETINOIC ACID-INDUCED GENE 1 PROTEIN AND
; FILE REFERENCE: 9800081-0085
; CURRENT APPLICATION NUMBER: US/10/224,289
; CURRENT FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: 60/313,940
; PRIOR FILING DATE: 2001-08-20
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1619
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-224-289-5

Query Match      100.0%; Score 42; DB 7; Length 1619;
Best Local Similarity 100.0%; Pred. No. 1e-07;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42
   |||||||
Db 1137 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1178

RESULT 16
US-10-935-190-43
; Sequence 43, Application US/10935190
; Publication No. US20050037466A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: AU-YOUNG, Janice
; APPLICANT: BANDMAN, Olga
; APPLICANT: TANG, Y. Tom
; APPLICANT: YUE, Henry
; APPLICANT: AZIMZAI, Yalda
; APPLICANT: BURFORD, Neil
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: LU, Dyrng Aina M.
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: PATTERSON, Chandra
; APPLICANT: LAL, Preeti
; TITLE OF INVENTION: RECEPTORS AND ASSOCIATED PROTEINS
; FILE REFERENCE: PF-0726 PCT
; CURRENT APPLICATION NUMBER: US/10/935,190
; CURRENT FILING DATE: 2004-09-08
; PRIOR APPLICATION NUMBER: US/10/031,904
; PRIOR FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: 60/145,232; 60/158,578; 60/165,192
; PRIOR FILING DATE: 1999-07-21; 1999-10-07; 1999-11-12
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PERL Program
; SEQ ID NO 43
```

```
; LENGTH: 1619
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 2681738CB1
US-10-935-190-43

Query Match      100.0%; Score 42; DB 9; Length 1619;
Best Local Similarity 100.0%; Pred. No. 1e-07;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42
   |||||||
Db 1137 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1178

RESULT 17
US-10-775-920-13
; Sequence 13, Application US/10775920
; Publication No. US20040175744A1
; GENERAL INFORMATION:
; APPLICANT: Mergen Ltd
; TITLE OF INVENTION: POLYNUCLEOTIDE SEQUENCES AND CORRESPONDING ENCODED POLYPEPTIDES
; TITLE OF INVENTION: OF PARTICULAR SECRETED AND MEMBRANE-BOUND PROTEINS OVEREXPRESSED
; FILE REFERENCE: Mergen - 0010B
; CURRENT APPLICATION NUMBER: US/10/775,920
; CURRENT FILING DATE: 2004-02-10
; PRIOR APPLICATION NUMBER: US 60/447,900
; PRIOR FILING DATE: 2003-02-13
; NUMBER OF SEQ ID NOS: 385
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 13
; LENGTH: 2297
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-775-920-13

Query Match      100.0%; Score 42; DB 9; Length 2297;
Best Local Similarity 100.0%; Pred. No. 1.1e-07;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42
   |||||||
Db 1125 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1166

RESULT 18
US-10-224-289-3
; Sequence 3, Application US/10224289
; Publication No. US20030207288A1
; GENERAL INFORMATION:
; APPLICANT: LEWIN, DAVID A.
; APPLICANT: STEWART, TIMOTHY A.
; TITLE OF INVENTION: GPCR-LIKE RETINOIC ACID-INDUCED GENE 1 PROTEIN AND
; TITLE OF INVENTION: NUCLEIC ACID
; FILE REFERENCE: 9800081-0085
; CURRENT APPLICATION NUMBER: US/10/224,289
; CURRENT FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: 60/313,940
; PRIOR FILING DATE: 2001-08-20
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 2302
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-224-289-3

Query Match      100.0%; Score 42; DB 7; Length 2302;
Best Local Similarity 100.0%; Pred. No. 1.1e-07;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy 1 GCCCAGCGTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
|||||  
Db 1117 GCCCAGCGTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 1158

## RESULT 19

US-10-425-405  
; Sequence 405, Application US/10240425  
; Publication No. US20040033502A1  
; GENERAL INFORMATION:  
; APPLICANT: Williams, Amanda  
; APPLICANT: Boland, Joseph F.  
; APPLICANT: Lord, Reginald V.  
; APPLICANT: Alvarez, Chris  
; APPLICANT: Wetzel, Jon C.  
; APPLICANT: Scherf, Uwe  
; APPLICANT: Vockley, Joseph G.  
; TITLE OF INVENTION: Gene Expression Profiles in Esophageal Tissue  
; FILE REFERENCE: 44921-5026  
; CURRENT APPLICATION NUMBER: US/10/240,425  
; CURRENT FILING DATE: 2002-09-30  
; PRIOR APPLICATION NUMBER: PCT/US01/09847  
; PRIOR FILING DATE: 2001-03-28  
; PRIOR APPLICATION NUMBER: US 60/193,446  
; PRIOR FILING DATE: 2000-03-31  
; NUMBER OF SEQ ID NOS: 1588  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 405  
; LENGTH: 2302  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: Genbank Accession No. US20040033502A1 AF0955448  
US-10-240-425-405

Query Match 100.0%; Score 42; DB 8; Length 2302;  
Best Local Similarity 100.0%; Pred. No. 1.1e-07;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
|||||  
Db 1117 GCCCAGCGTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 1158

## RESULT 20

US-10-775-920-9  
; Sequence 9, Application US/10775920  
; Publication No. US2004017544A1  
; GENERAL INFORMATION:  
; APPLICANT: Merzen Ltd  
; TITLE OF INVENTION: POLYNUCLEOTIDE SEQUENCES AND CORRESPONDING ENCODED POLYPEPTIDES  
; TITLE OF INVENTION: OF PARTICULAR SECRETED AND MEMBRANE-BOUND PROTEINS OVEREXPRESSED  
; FILE REFERENCE: Merzen - 0010B  
; CURRENT APPLICATION NUMBER: US/10/775,920  
; CURRENT FILING DATE: 2004-02-10  
; PRIOR APPLICATION NUMBER: US 60/447,900  
; PRIOR FILING DATE: 2003-02-13  
; NUMBER OF SEQ ID NOS: 385  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 9  
; LENGTH: 2302  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-775-920-9

Query Match 100.0%; Score 42; DB 9; Length 2302;  
Best Local Similarity 100.0%; Pred. No. 1.1e-07;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
|||||

Db 1117 GCCCAGCGTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 1158

## RESULT 21

US-10-510-507-2  
; Sequence 2, Application US/10510507  
; Publication No. US20050282165A1  
; GENERAL INFORMATION:  
; APPLICANT: Terrett, Jonathan A  
; TITLE OF INVENTION: DIAGNOSIS OF CARCINOMA USING RAIG1 POLYPEPTIDES  
; FILE REFERENCE: 2543-1-039PCT/US  
; CURRENT APPLICATION NUMBER: US/10/510,507  
; CURRENT FILING DATE: 2004-10-07  
; PRIOR APPLICATION NUMBER: GB0208331.9  
; PRIOR FILING DATE: 2002-04-11  
; PRIOR APPLICATION NUMBER: GB0221538.2  
; PRIOR FILING DATE: 2002-09-17  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 2302  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-510-507-2

Query Match 100.0%; Score 42; DB 10; Length 2302;  
Best Local Similarity 100.0%; Pred. No. 1.1e-07;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
|||||  
Db 1117 GCCCAGCGTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 1158

## RESULT 22

US-10-775-920-12  
; Sequence 12, Application US/10775920  
; Publication No. US2004017544A1  
; GENERAL INFORMATION:  
; APPLICANT: Merzen Ltd  
; TITLE OF INVENTION: POLYNUCLEOTIDE SEQUENCES AND CORRESPONDING ENCODED POLYPEPTIDES  
; TITLE OF INVENTION: OF PARTICULAR SECRETED AND MEMBRANE-BOUND PROTEINS OVEREXPRESSED  
; FILE REFERENCE: Merzen - 0010B  
; CURRENT APPLICATION NUMBER: US/10/775,920  
; CURRENT FILING DATE: 2004-02-10  
; PRIOR APPLICATION NUMBER: US 60/447,900  
; PRIOR FILING DATE: 2003-02-13  
; NUMBER OF SEQ ID NOS: 385  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 12  
; LENGTH: 2305  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-775-920-12

Query Match 100.0%; Score 42; DB 9; Length 2305;  
Best Local Similarity 100.0%; Pred. No. 1.1e-07;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
|||||  
Db 1121 GCCCAGCGTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 1162

## RESULT 23

US-10-176-847-59  
; Sequence 59, Application US/10176847  
; Publication No. US20030068636A1  
; GENERAL INFORMATION:  
; APPLICANT: Veiby, Petter Ole  
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR  
; IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF BREAST

```
; TITLE OF INVENTION: AND OVARIAN CANCER
; FILE REFERENCE: MRI-039
; CURRENT APPLICATION NUMBER: US/10/176,847
; CURRENT FILING DATE: 2002-06-21
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 59
; LENGTH: 2316
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-176-847-59

Query Match 100.0%; Score 42; DB 6; Length 2316;
Best Local Similarity 100.0%; Pred. No. 1.1e-07;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42
Db 1123 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1164

RESULT 24
US-11-080-991-59
; Sequence 59, Application US/11080991
; Publication No. US20050266437A1
; GENERAL INFORMATION:
; APPLICANT: Veiby, Pette Ole
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF BREAST
; TITLE OF INVENTION: AND OVARIAN CANCER
; FILE REFERENCE: MRI-039
; CURRENT APPLICATION NUMBER: US/11/080,991
; CURRENT FILING DATE: 2005-03-11
; PRIOR APPLICATION NUMBER: US/10/176,847
; PRIOR FILING DATE: 2002-06-21
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 59
; LENGTH: 2316
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-080-991-59

Query Match 100.0%; Score 42; DB 15; Length 2316;
Best Local Similarity 100.0%; Pred. No. 1.1e-07;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42
Db 1123 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1164

RESULT 25
US-10-775-920-11
; Sequence 11, Application US/10775920
; Publication No. US2004017544A1
; GENERAL INFORMATION:
; APPLICANT: Merigen Ltd
; TITLE OF INVENTION: POLYNUCLEOTIDE SEQUENCES AND CORRESPONDING ENCODED POLYPEPTIDES
; TITLE OF INVENTION: OF PARTICULAR SECRETED AND MEMBRANE-BOUND PROTEINS OVEREXPRESSED
; TITLE OF INVENTION: IN CERTAIN CANCERS
; FILE REFERENCE: Merigen - 00108
; CURRENT APPLICATION NUMBER: US/10/775,920
; CURRENT FILING DATE: 2004-02-10
; PRIOR APPLICATION NUMBER: US 60/447,900
; PRIOR FILING DATE: 2003-02-13
; NUMBER OF SEQ ID NOS: 385
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11
; LENGTH: 2446
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-775-920-11
```

```
Query Match 100.0%; Score 42; DB 9; Length 2446;
Best Local Similarity 100.0%; Pred. No. 1.1e-07;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42
Db 1271 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1312

RESULT 26
US-10-225-567A-453
; Sequence 453, Application US/10225567A
; Publication No. US20030113798A1
; GENERAL INFORMATION:
; APPLICANT: LifeSpan Biosciences
; APPLICANT: Brown, Joseph P.
; APPLICANT: Burner, Glenna C.
; APPLICANT: Roush, Christine L.
; TITLE OF INVENTION: ANTIGENIC PEPTIDES AND ANTIBODIES FOR G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 1920-4-4
; CURRENT APPLICATION NUMBER: US/10/225,567A
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/257,144
; PRIOR FILING DATE: 2000-12-19
; NUMBER OF SEQ ID NOS: 2292
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 453
; LENGTH: 2456
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-225-567A-453

Query Match 100.0%; Score 42; DB 6; Length 2456;
Best Local Similarity 100.0%; Pred. No. 1.1e-07;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42
Db 1271 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1312

RESULT 27
US-10-269-909-63
; Sequence 63, Application US/10269909
; Publication No. US20030180747A1
; GENERAL INFORMATION:
; APPLICANT: HRUBAN, RALPH H.
; APPLICANT: ARGANT, PEDRAM
; APPLICANT: IACOBUZIO-DONAHUE, CHRISTINE
; APPLICANT: MAITRA, ANIRBAN
; TITLE OF INVENTION: PANCREATIC CANCER DIAGNOSIS AND THERAPIES
; FILE REFERENCE: 58303(71699)
; CURRENT APPLICATION NUMBER: US/10/269,909
; CURRENT FILING DATE: 2003-10-11
; PRIOR APPLICATION NUMBER: 60/328,609
; PRIOR FILING DATE: 2001-10-11
; PRIOR APPLICATION NUMBER: 60/332,754
; PRIOR FILING DATE: 2001-11-19
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 63
; LENGTH: 2456
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-269-909-63

Query Match 100.0%; Score 42; DB 7; Length 2456;
Best Local Similarity 100.0%; Pred. No. 1.1e-07;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42
Db 1123 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1164
```

Db 1271 GCCCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1312

## RESULT 28

US-10-269-909-64  
; Sequence 64, Application US/10269909  
; Publication No. US20030180747A1  
; GENERAL INFORMATION:  
; APPLICANT: HRUBAN, RALPH H.  
; APPLICANT: ARGANI, PEDRAM  
; APPLICANT: IACOBUIZIO-DONAHUE, CHRISTINE  
; APPLICANT: MAITRA, ANIRBAN  
; TITLE OF INVENTION: PANCREATIC CANCER DIAGNOSIS AND THERAPIES  
; FILE REFERENCE: 58303(71699)  
; CURRENT APPLICATION NUMBER: US/10/269,909  
; CURRENT FILING DATE: 2003-10-11  
; PRIOR APPLICATION NUMBER: 60/328,609  
; PRIOR FILING DATE: 2001-10-11  
; PRIOR APPLICATION NUMBER: 60/332,754  
; PRIOR FILING DATE: 2001-11-19  
; NUMBER OF SEQ ID NOS: 87  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 64  
; LENGTH: 2456  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-269-909-64

Query Match 100.0%; Score 42; DB 7; Length 2456;  
Best Local Similarity 100.0%; Pred. No. 1.1e-07;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42

Db 1271 GCCCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1312

## RESULT 29

US-10-295-027-619  
; Sequence 619, Application US/10295027  
; Publication No. US2003023250A1  
; GENERAL INFORMATION:  
; APPLICANT: Afar, Daniel  
; APPLICANT: Aziz, Natasha  
; APPLICANT: Ginsberg, Wendy M.  
; APPLICANT: Gish, Kurt C.  
; APPLICANT: Glynn, Richard  
; APPLICANT: Hevezi, Peter A.  
; APPLICANT: Mack, David H.  
; APPLICANT: Murray, Richard  
; APPLICANT: Watson, Susan R.  
; APPLICANT: Eos Biotechnology, Inc.  
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and  
; FILE REFERENCE: 018501-01250005  
; CURRENT APPLICATION NUMBER: US/10/295,027  
; CURRENT FILING DATE: 2002-11-13  
; PRIOR APPLICATION NUMBER: US 09/663,733  
; PRIOR FILING DATE: 2000-09-15  
; PRIOR APPLICATION NUMBER: US 60/350,666  
; PRIOR FILING DATE: 2001-11-13  
; PRIOR APPLICATION NUMBER: US 60/335,394  
; PRIOR FILING DATE: 2001-11-15  
; PRIOR APPLICATION NUMBER: US 60/332,464  
; PRIOR FILING DATE: 2001-11-21  
; PRIOR APPLICATION NUMBER: US 60/334,393  
; PRIOR FILING DATE: 2001-11-29  
; PRIOR APPLICATION NUMBER: US 60/340,376  
; PRIOR FILING DATE: 2001-12-14  
; PRIOR APPLICATION NUMBER: US 60/347,211  
; PRIOR FILING DATE: 2002-01-08  
; PRIOR APPLICATION NUMBER: US 60/347,349  
; PRIOR FILING DATE: 2002-01-10

; PRIOR APPLICATION NUMBER: US 60/355,250  
; PRIOR FILING DATE: 2002-02-08  
; PRIOR APPLICATION NUMBER: US 60/356,714  
; PRIOR FILING DATE: 2002-02-13  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 1386  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 619  
; LENGTH: 2456  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-295-027-619

Query Match 100.0%; Score 42; DB 7; Length 2456;  
Best Local Similarity 100.0%; Pred. No. 1.1e-07;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42

Db 1271 GCCCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1312

## RESULT 30

US-10-600-816-2  
; Sequence 2, Application US/10600816  
; Publication No. US20040121362A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: IDENTIFICATION AND MODULATION OF A G-PROTEIN COUPLED RECEPTOR  
; FILE REFERENCE: D0251 NP  
; CURRENT APPLICATION NUMBER: US/10/600,816  
; CURRENT FILING DATE: 2003-06-20  
; PRIOR APPLICATION NUMBER: U.S. 60/390,850  
; PRIOR FILING DATE: 2002-06-20  
; PRIOR APPLICATION NUMBER: U.S. 60/407,006  
; PRIOR FILING DATE: 2002-08-29  
; NUMBER OF SEQ ID NOS: 98  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 2  
; LENGTH: 2456  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-600-816-2

Query Match 100.0%; Score 42; DB 8; Length 2456;  
Best Local Similarity 100.0%; Pred. No. 1.1e-07;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42

Db 1271 GCCCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1312

## RESULT 31

US-10-600-816-18  
; Sequence 18, Application US/10600816  
; Publication No. US20040121362A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: IDENTIFICATION AND MODULATION OF A G-PROTEIN COUPLED RECEPTOR  
; FILE REFERENCE: D0251 NP  
; CURRENT APPLICATION NUMBER: US/10/600,816  
; CURRENT FILING DATE: 2003-06-20  
; PRIOR APPLICATION NUMBER: U.S. 60/390,850  
; PRIOR FILING DATE: 2002-06-20  
; PRIOR APPLICATION NUMBER: U.S. 60/407,006  
; PRIOR FILING DATE: 2002-08-29  
; NUMBER OF SEQ ID NOS: 98  
; SOFTWARE: PatentIn version 3.2

```
/ SEQ ID NO 18
/ LENGTH: 2456
/ TYPE: DNA
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: RA13 Polymorphic Allele Summary Sequence.
/
/ NAME/KEY: misc_feature
/ LOCATION: (112)..(112)
/ OTHER INFORMATION: wherein "n" equals either G or A.
/
/ NAME/KEY: misc_feature
/ LOCATION: (364)..(364)
/ OTHER INFORMATION: wherein "n" equals either C or T.
/
/ NAME/KEY: misc_feature
/ LOCATION: (511)..(511)
/ OTHER INFORMATION: wherein "n" equals either C or T.
/
/ NAME/KEY: misc_feature
/ LOCATION: (523)..(523)
/ OTHER INFORMATION: wherein "n" equals either C or T.
/
/ NAME/KEY: misc_feature
/ LOCATION: (605)..(605)
/ OTHER INFORMATION: wherein "n" equals either A or G.
/
/ NAME/KEY: misc_feature
/ LOCATION: (797)..(797)
/ OTHER INFORMATION: wherein "n" equals either A or G.
/
/ NAME/KEY: misc_feature
/ LOCATION: (1111)..(1111)
/ OTHER INFORMATION: wherein "n" equals either T or C.
/
/ NAME/KEY: misc_feature
/ LOCATION: (1173)..(1173)
/ OTHER INFORMATION: wherein "n" equals either A or G.
/
/
/ US-10-600-816-18
/
/ Query Match
/ Best Local Similarity 100.0%; Score 42; DB 8; Length 2456;
/ Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
/
/ QY 1 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42
/ Db 1271 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1312
/
/ RESULT 32
/ US-10-775-920-10
/ Sequence 10, Application US/10775920
/ Publication No. US20040175744A1
/ GENERAL INFORMATION:
/ APPLICANT: Mergen Ltd
/ TITLE OF INVENTION: POLYNUCLEOTIDE SEQUENCES AND CORRESPONDING ENCODED POLYPEPTIDES
/ FILE REFERENCE: IN CERTAIN CANCERS
/ CURRENT APPLICATION NUMBER: US/10/775,920
/ PRIOR FILING DATE: 2004-02-10
/ NUMBER OF SEQ ID NOS: 385
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 10
/ LENGTH: 2456
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ US-10-775-920-10
/
/ Query Match
/ Best Local Similarity 100.0%; Score 42; DB 9; Length 2456;
/ Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
/
/
/ Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
/
/ QY 1 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42
/ Db 1271 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1312
/
/ RESULT 33
/ US-10-936-626-40
/ Sequence 40, Application US/10936626
/ Publication No. US20050106644A1
/ GENERAL INFORMATION:
/ APPLICANT: Cairns, Belinda
/ APPLICANT: Chen, Ruihuan
/ APPLICANT: Frantz, Gretchen
/ APPLICANT: Hillan, Kenneth J.
/ APPLICANT: Koepfen, Hartmut
/ APPLICANT: Phillips, Heidi S.
/ APPLICANT: Polakis, Paul
/ APPLICANT: Spencer, Susan D.
/ APPLICANT: Smith, Victoria
/ APPLICANT: Williams, P. Mickey
/ APPLICANT: Wu, Thomas D.
/ APPLICANT: Zhang, Zemin
/ TITLE OF INVENTION: Compositions and Methods for the Diagnosis and
/ TITLE OF INVENTION: Treatment of Tumor
/ FILE REFERENCE: P5001R1P1
/ CURRENT APPLICATION NUMBER: US/10/936,626
/ CURRENT FILING DATE: 2004-09-08
/ PRIOR APPLICATION NUMBER: US 10/872,991
/ PRIOR FILING DATE: 2004-06-21
/ PRIOR APPLICATION NUMBER: US 10/872,972
/ PRIOR FILING DATE: 2004-06-21
/ PRIOR APPLICATION NUMBER: US 10/241,220
/ PRIOR FILING DATE: 2002-09-11
/ PRIOR APPLICATION NUMBER: US 10/177,488
/ PRIOR FILING DATE: 2002-06-19
/ PRIOR APPLICATION NUMBER: US 60/299,500
/ PRIOR FILING DATE: 2001-06-20
/ PRIOR APPLICATION NUMBER: US 60/301,880
/ PRIOR FILING DATE: 2001-06-29
/ PRIOR APPLICATION NUMBER: US 60/323,268
/ PRIOR FILING DATE: 2001-09-18
/ PRIOR APPLICATION NUMBER: US 60/557,116
/ PRIOR FILING DATE: 2004-03-26
/ PRIOR APPLICATION NUMBER: US 60/598,899
/ PRIOR FILING DATE: 2004-08-04
/ NUMBER OF SEQ ID NOS: 154
/ SEQ ID NO 40
/ LENGTH: 2456
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ US-10-936-626-40
/
/ Query Match
/ Best Local Similarity 100.0%; Score 42; DB 10; Length 2456;
/ Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
/
/ QY 1 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42
/ Db 1271 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1312
/
/ RESULT 34
/ US-10-938-061-40
/ Sequence 40, Application US/10938061
/ Publication No. US20050107595A1
/ GENERAL INFORMATION:
/ APPLICANT: Cairns, Belinda
/ APPLICANT: Chen, Ruihuan
/ APPLICANT: Frantz, Gretchen
/ APPLICANT: Hillan, Kenneth J.
/ APPLICANT: Koepfen, Hartmut
```



APPLICANT: Phillips, Heidi S.  
APPLICANT: Polakis, Paul  
APPLICANT: Spencer, Susan D.  
APPLICANT: Smith, Victoria  
APPLICANT: Williams, P. Mickey  
APPLICANT: Wu, Thomas D.  
APPLICANT: Zhang, Zemin  
APPLICANT: Sakanaka, Chile  
APPLICANT: Chuntharapai, Anan  
APPLICANT: Reed Chae J.  
TITLE OF INVENTION: Compositions and Methods for the Diagnosis and  
TITLE OF INVENTION: Treatment of Tumor  
FILE REFERENCE: P5001R1PIB  
CURRENT APPLICATION NUMBER: US/10/938,061  
CURRENT FILING DATE: 2004-09-10  
PRIOR APPLICATION NUMBER: US 10/872,991  
PRIOR FILING DATE: 2004-06-21  
PRIOR APPLICATION NUMBER: US 10/872,972  
PRIOR FILING DATE: 2004-06-21  
PRIOR APPLICATION NUMBER: US 10/241,220  
PRIOR FILING DATE: 2002-09-11  
PRIOR APPLICATION NUMBER: US 10/177,488  
PRIOR FILING DATE: 2002-06-19  
PRIOR APPLICATION NUMBER: US 60/299,500  
PRIOR FILING DATE: 2001-06-20  
PRIOR APPLICATION NUMBER: US 60/301,880  
PRIOR FILING DATE: 2001-06-29  
PRIOR APPLICATION NUMBER: US 60/323,268  
PRIOR FILING DATE: 2001-09-18  
PRIOR APPLICATION NUMBER: US 60/557,116  
PRIOR FILING DATE: 2004-03-26  
PRIOR APPLICATION NUMBER: US 60/598,899  
PRIOR FILING DATE: 2004-08-04  
NUMBER OF SEQ ID NOS: 154  
SEQ ID NO 40  
LENGTH: 2456  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-938-061-40

Query Match 100.0%; Score 42; DB 10; Length 2456;  
Best Local Similarity 100.0%; Pred. No. 1.1e-07;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
|||||  
Db 1271 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 1312

RESULT 35  
US-11-169-041-32  
Sequence 32, Application US/11169041  
Publication No. US20060019284A1  
GENERAL INFORMATION:  
APPLICANT: Bristol-Myers Squibb Company  
TITLE OF INVENTION: IDENTIFICATION OF POLYNUCLEOTIDES FOR PREDICTING ACTIVITY OF  
TITLE OF INVENTION: COMPOUNDS THAT INTERACT WITH AND/OR MODULATE PROTEIN TYROSINE  
TITLE OF INVENTION: KINASES AND/OR PROTEIN TYROSINE KINASE PATHWAYS IN LUNG CANCER  
TITLE OF INVENTION: CELLS  
FILE REFERENCE: 10001 NP  
CURRENT APPLICATION NUMBER: US/11/169,041  
CURRENT FILING DATE: 2005-06-28  
PRIOR APPLICATION NUMBER: 60/584,405  
PRIOR FILING DATE: 2004-06-30  
NUMBER OF SEQ ID NOS: 527  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 32  
LENGTH: 2456  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-11-169-041-32

Query Match 100.0%; Score 42; DB 16; Length 2456;

Best Local Similarity 100.0%; Pred. No. 1.1e-07;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
|||||  
Db 1271 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 1312

RESULT 36  
US-10-264-049-834  
Sequence 834, Application US/10264049  
Publication No. US20040005579A1  
GENERAL INFORMATION:  
APPLICANT: Birse et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PA133P1  
CURRENT APPLICATION NUMBER: US/10/264,049  
CURRENT FILING DATE: 2002-10-04  
PRIOR APPLICATION NUMBER: PCT/US01/18569  
PRIOR FILING DATE: 2001-06-07  
PRIOR APPLICATION NUMBER: US 60/209,467  
PRIOR FILING DATE: 2000-06-07  
NUMBER OF SEQ ID NOS: 4360  
SOFTWARE: PatentIn Ver. 3.1  
SEQ ID NO 834  
LENGTH: 2593  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-264-049-834

Query Match 100.0%; Score 42; DB 7; Length 2593;  
Best Local Similarity 100.0%; Pred. No. 1.2e-07;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
|||||  
Db 1383 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 1424

RESULT 37  
US-10-198-846-10424  
Sequence 10424, Application US/10198846  
Publication No. US20030099974A1  
GENERAL INFORMATION:  
APPLICANT: Lillie, James  
APPLICANT: Xu, Yongyao  
APPLICANT: Wang, Youzhen  
APPLICANT: Steinmann, Kathleen  
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS  
TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
TITLE OF INVENTION: THERAPY OF BREAST CANCER  
FILE REFERENCE: MRI-049  
CURRENT APPLICATION NUMBER: US/10/198,846  
CURRENT FILING DATE: 2002-07-18  
PRIOR APPLICATION NUMBER: 60/306,220  
PRIOR FILING DATE: 2001-07-18  
NUMBER OF SEQ ID NOS: 14084  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 10424  
LENGTH: 4239  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 1, 4238, 4239  
OTHER INFORMATION: n = A,T,C or G  
US-10-198-846-10424

Query Match 100.0%; Score 42; DB 6; Length 4239;  
Best Local Similarity 100.0%; Pred. No. 1.3e-07;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 42

Db 1679 GCCCAGCGTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 1720  
|||||  
RESULT 38  
US-10-505-486-196  
; Sequence 196, Application US/10505486  
; Publication No. US20050118639A1  
; GENERAL INFORMATION:  
; APPLICANT: Takeda Chemical Industries, Ltd.  
; TITLE OF INVENTION: Determination of a ligand  
; FILE REFERENCE: P03-0008PCT  
; CURRENT APPLICATION NUMBER: US/10/505,486  
; PRIOR FILING DATE: 2004-08-20  
; PRIOR APPLICATION NUMBER: JP 2002-45728  
; PRIOR FILING DATE: 2002-02-22  
; PRIOR APPLICATION NUMBER: JP 2002-213949  
; PRIOR FILING DATE: 2002-07-23  
; PRIOR APPLICATION NUMBER: JP 2002-298237  
; PRIOR FILING DATE: 2002-10-11  
; NUMBER OF SEQ ID NOS: 233  
; SEQ ID NO 196  
; LENGTH: 1788  
; TYPE: DNA  
; ORGANISM: Human  
US-10-505-486-196

Query Match 96.2%; Score 40.4; DB 10; Length 1788;  
Best Local Similarity 97.6%; Pred. No. 5.3e-07;  
Matches 41; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCCAGCGTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
|||||  
Db 1018 GCCCAGCGTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 1059

RESULT 39  
US-10-600-816-33  
; Sequence 33, Application US/10600816  
; Publication No. US20040121362A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: IDENTIFICATION AND MODULATION OF A G-PROTEIN COUPLED RECEPTOR  
; TITLE OF INVENTION: (GPCR), RA13, ASSOCIATED WITH CHRONIC OBSTRUCTIVE PULMONARY  
; TITLE OF INVENTION: DISEASE (COPD) AND NF-KB AND E-SELECTIN REGULATION  
; FILE REFERENCE: D0251 NP  
; CURRENT APPLICATION NUMBER: US/10/600,816  
; CURRENT FILING DATE: 2003-06-20  
; PRIOR APPLICATION NUMBER: U.S. 60/390,850  
; PRIOR FILING DATE: 2002-06-20  
; PRIOR APPLICATION NUMBER: U.S. 60/407,006  
; PRIOR FILING DATE: 2002-08-29  
; NUMBER OF SEQ ID NOS: 98  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 33  
; LENGTH: 42  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Degenerate Oligonucleotide.  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (3)..(39)  
; OTHER INFORMATION: wherein "n" equals A, C, G, or T.  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (6)..(33)  
; OTHER INFORMATION: wherein "y" equals C, or T.  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (27)..(42)  
; OTHER INFORMATION: wherein "r" equals A, or G.  
US-10-600-816-33

Query Match 71.4%; Score 30; DB 8; Length 42;  
Best Local Similarity 64.3%; Pred. No. 0.0081;  
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 GCCCAGCGTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
|||||  
Db 1 GCNCAYCMTGGCCNTCNCNTAYAAAGATTAYGARGTNAAR 42  
|||||

RESULT 40  
US-11-060-756-43361  
; Sequence 43361, Application US/11060756  
; Publication No. US20050221354A1  
; GENERAL INFORMATION:  
; APPLICANT: Wyeth  
; APPLICANT: Mounts, William Martin  
; TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug  
; TITLE OF INVENTION: Target Genes  
; FILE REFERENCE: AM101083 (031896-042000)  
; CURRENT APPLICATION NUMBER: US/11/060,756  
; CURRENT FILING DATE: 2005-02-18  
; NUMBER OF SEQ ID NOS: 303284  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 43361  
; LENGTH: 25  
; TYPE: DNA  
; ORGANISM: probe  
US-11-060-756-43361

Query Match 59.5%; Score 25; DB 13; Length 25;  
Best Local Similarity 100.0%; Pred. No. 1.1;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 GAGCCCTTACAAAGACTATGAAGTA 39  
|||||  
Db 1 GAGCCCTTACAAAGACTATGAAGTA 25  
|||||

RESULT 41  
US-10-459-155-7  
; Sequence 7, Application US/10459155  
; Publication No. US20040254365A1  
; GENERAL INFORMATION:  
; APPLICANT: Lee, Sho Tone  
; TITLE OF INVENTION: IMMUNIZATION AGAINST FLAVIVIRUS  
; FILE REFERENCE: 08919-101001  
; CURRENT APPLICATION NUMBER: US/10/459,155  
; CURRENT FILING DATE: 2003-06-11  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7  
; LENGTH: 480  
; TYPE: DNA  
; ORGANISM: Japanese encephalitis virus  
US-10-459-155-7

Query Match 54.3%; Score 22.8; DB 9; Length 480;  
Best Local Similarity 79.4%; Pred. No. 21;  
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 CAGCGTTGGCGAGCCCTTACAAAGACTATGAAG 37  
|||||  
Db 330 CAGCGTTGGCGAGCCCTTTTCAACGACTTTGAAG 363  
|||||

RESULT 42  
US-10-459-155-12  
; Sequence 12, Application US/10459155  
; Publication No. US20040254365A1  
; GENERAL INFORMATION:  
; APPLICANT: Lee, Sho Tone  
; TITLE OF INVENTION: IMMUNIZATION AGAINST FLAVIVIRUS

```
; FILE REFERENCE: 08919-101001
; CURRENT APPLICATION NUMBER: US/10/459.155
; CURRENT FILING DATE: 2003-06-11
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 525
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetically generated oligonucleotide
US-10-459-155-12

Query Match          54.3%; Score 22.8; DB 9; Length 525;
Best Local Similarity 79.4%; Pred. No. 21;
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy  4 CACGCTTGGCGAGCCCTTACAAAGACTATGAAG 37
      ||||| ||| ||||| ||| ||||| |||||
Db  375 CACGCTGGGCAAGCCCTTTTCAACGACTTTGAAG 408

RESULT 43
US-10-459-155-9
; Sequence 9, Application US/10459155
; Publication No. US20040254365A1
; GENERAL INFORMATION:
; APPLICANT: Lee, Sho Tone
; TITLE OF INVENTION: IMMUNIZATION AGAINST FLAVIVIRUS
; FILE REFERENCE: 08919-101001
; CURRENT APPLICATION NUMBER: US/10/459.155
; CURRENT FILING DATE: 2003-06-11
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 1500
; TYPE: DNA
; ORGANISM: Japanese encephalitis virus
US-10-459-155-9

Query Match          54.3%; Score 22.8; DB 9; Length 1500;
Best Local Similarity 79.4%; Pred. No. 28;
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy  4 CACGCTTGGCGAGCCCTTACAAAGACTATGAAG 37
      ||||| ||| ||||| ||| ||||| |||||
Db  1203 CACGCTGGGCAAGCCCTTTTCAACGACTTTGAAG 1236

RESULT 44
US-10-459-155-21
; Sequence 21, Application US/10459155
; Publication No. US20040254365A1
; GENERAL INFORMATION:
; APPLICANT: Lee, Sho Tone
; TITLE OF INVENTION: IMMUNIZATION AGAINST FLAVIVIRUS
; FILE REFERENCE: 08919-101001
; CURRENT APPLICATION NUMBER: US/10/459.155
; CURRENT FILING DATE: 2003-06-11
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 1545
; TYPE: DNA
; ORGANISM: Japanese encephalitis virus
US-10-459-155-21

Query Match          54.3%; Score 22.8; DB 9; Length 1545;
Best Local Similarity 79.4%; Pred. No. 28;
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy  4 CACGCTTGGCGAGCCCTTACAAAGACTATGAAG 37
      ||||| ||| ||||| ||| ||||| |||||
```

```
Db  1248 CACGCTGGCAAGCCCTTTTCAACGACTTTGAAG 1281

RESULT 45
US-09-925-065A-691738
; Sequence 691738, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925.065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 691738
; LENGTH: 1540
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-691738

Query Match          53.3%; Score 22.4; DB 4; Length 1540;
Best Local Similarity 72.5%; Pred. No. 41;
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy  3 CCAGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 42
      ||||| ||| ||||| ||| ||||| |||||
Db  870 CAAACTAGACCAAGGCTTTACAAAAAACTAAGAAGTATAG 909

RESULT 46
US-09-925-065A-691739
; Sequence 691739, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925.065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 691739
; LENGTH: 1540
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-691739

Query Match          53.3%; Score 22.4; DB 4; Length 1540;
Best Local Similarity 72.5%; Pred. No. 41;
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
```

```

Query Match      53.3%; Score 22.4; DB 4; Length 1540;
Best Local Similarity 72.5%; Pred. No. 41;
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY      3 CCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42
      |||||
Db      870 CAAAACTAGACCAAGGCTTTACAAAACCTAAGAAGTATAG 909

RESULT 47
US-09-925-065A-691740
; Sequence 691740, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 691740
; LENGTH: 1540
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-691740

Query Match      53.3%; Score 22.4; DB 4; Length 1540;
Best Local Similarity 72.5%; Pred. No. 41;
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY      3 CCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42
      |||||
Db      870 CAAAACTAGACCAAGGCTTTACAAAACCTAAGAAGTATAG 909

RESULT 48
US-09-925-065A-691738
; Sequence 691738, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 691738
; LENGTH: 1540
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-691738

Query Match      53.3%; Score 22.4; DB 5; Length 1540;
Best Local Similarity 72.5%; Pred. No. 41;
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY      3 CCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42
      |||||
Db      870 CAAAACTAGACCAAGGCTTTACAAAACCTAAGAAGTATAG 909

RESULT 50
US-09-925-065A-691740
; Sequence 691740, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 691740
; LENGTH: 1540
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-691739

Query Match      53.3%; Score 22.4; DB 5; Length 1540;
Best Local Similarity 72.5%; Pred. No. 41;
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY      3 CCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42
      |||||
Db      870 CAAAACTAGACCAAGGCTTTACAAAACCTAAGAAGTATAG 909

RESULT 51
US-09-925-065A-691740
; Sequence 691740, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 691740
; LENGTH: 1540
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-691739
```

```
; ORGANISM: Homo sapiens
US-09-925-065A-691740

Query Match      53.3%; Score 22.4; DB 5; Length 1540;
Best Local Similarity 72.5%; Pred. No. 41;
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy      3 CCAGCTTGCGGAGCCCTTACAAAGACTATGAAGTAAAG 42
      |||
Db      870 CAAACTAGACCAAGGCTTTACAAAGAACTAAGAAGTATAG 909

RESULT 51
US-09-917-800A-76
; Sequence 76, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Castle, Arthur
; APPLICANT: Elashoff, Michael
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917,800A
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/222,880
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/290,029
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/290,645
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: US 60/292,336
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/295,798
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/297,457
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,884
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,459
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 1740
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 76
; LENGTH: 584
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020119462A1 AA818700
US-09-917-800A-76

Query Match      52.9%; Score 22.2; DB 3; Length 584;
Best Local Similarity 88.9%; Pred. No. 40;
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      15 GAGCCCTTACAAAGACTATGAAGTAAA 41
      |||
Db      373 GAGCCCTTACAAAGACTATGAAGACCAA 399

RESULT 52
US-09-925-065A-49510
; Sequence 49510, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135

; ORGANISM: Homo sapiens
US-09-925-065A-49510

Query Match      52.9%; Score 22.2; DB 5; Length 2548;
Best Local Similarity 77.1%; Pred. No. 57;
Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy      7 GCCTTGCCGAGCCCTTACAAAGACTATGAAGTAAA 41
      |||
Db      374 GCCTTGACAAGTTATTATCAAAAGATATGAAGTAAA 408

RESULT 53
US-09-925-065A-49510
; Sequence 49510, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 49510
; LENGTH: 2548
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-49510

Query Match      52.9%; Score 22.2; DB 5; Length 2548;
Best Local Similarity 77.1%; Pred. No. 57;
Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy      7 GCCTTGCCGAGCCCTTACAAAGACTATGAAGTAAA 41
      |||
Db      374 GCCTTGACAAGTTATTATCAAAAGATATGAAGTAAA 408

RESULT 54
US-10-301-480-150748
; Sequence 150748, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
```

```
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms
; TITLE OF INVENTION: in the Human Genome
; FILE REFERENCE: 10827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 150748
; LENGTH: 2548
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-150748

Query Match          52.9%; Score 22.2; DB 12; Length 2548;
Best Local Similarity 77.1%; Pred. No. 57;
Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY      7 GCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAA 41
Db      374 GCCTTGACAAGTTATTACAAAGATATGAAGTAAA 408

RESULT 55
US-10-301-480-764157
; Sequence 764157, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms
; FILE REFERENCE: 10827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 764157
; LENGTH: 2548
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-764157

Query Match          52.9%; Score 22.2; DB 12; Length 2548;
Best Local Similarity 77.1%; Pred. No. 57;
Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY      7 GCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAA 41
Db      374 GCCTTGACAAGTTATTACAAAGATATGAAGTAAA 408

RESULT 56
US-10-017-161-455/c
; Sequence 455, Application US/10017161
; Publication No. US20030143669A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
```

```
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 455
; LENGTH: 1339
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: source
; LOCATION: (1)..(1339)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(1139)
US-10-017-161-455
```

```
Query Match          51.9%; Score 21.8; DB 7; Length 1339;
Best Local Similarity 78.8%; Pred. No. 74;
Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
```

```
QY      4 CACGCTTGCCGAGCCCTTACAAAGACTATGAA 36
Db      91 CACCCTTCCTGGAGTCCTTACAAAGACAATGAA 59
```

```
RESULT 57
US-10-292-798-409/c
; Sequence 409, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
; FILE REFERENCE: 084335/166
; CURRENT APPLICATION NUMBER: US/10/292,798
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: 10/017,161
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: JP 2001-246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2070
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 409
; LENGTH: 1339
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; LOCATION: source
; FEATURE:
; LOCATION: (1)..(1339)
; NAME/KEY: CDS
; LOCATION: (201)..(1139)
US-10-292-798-409
```

```
Query Match          51.9%; Score 21.8; DB 7; Length 1339;
Best Local Similarity 78.8%; Pred. No. 74;
Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
```

```
QY      4 CACGCTTGCCGAGCCCTTACAAAGACTATGAA 36
Db      91 CACCCTTCCTGGAGTCCTTACAAAGACAATGAA 59
```

```
RESULT 58
US-10-044-090-126/c
; Sequence 126, Application US/10044090
; Publication No. US20020137081A1
; GENERAL INFORMATION:
; APPLICANT: Olga Bandman
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
; FILE REFERENCE: PA-0028 US
; CURRENT APPLICATION NUMBER: US/10/044,090
```

```
; CURRENT FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 126
; LENGTH: 4826
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20020137081A1 238854.23
US-10-044-090-126

Query Match          51.9%; Score 21.8; DB 6; Length 4826;
Best Local Similarity 70.7%; Pred. No. 1e+02;
Matches 29; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAA 41
Db 4091 GCGAAGTTTCACAGAGTCTTCAAAAGACTCTGAAGAAA 4051

RESULT 59
US-10-084-817-304/c
; Sequence 304, Application US/10084817
; Publication No. US20030119009A1
; GENERAL INFORMATION:
; APPLICANT: Susan Stuart
; APPLICANT: Jed G. Nuchtern
; APPLICANT: Sharon E. Pion
; APPLICANT: Jason M. Shohet
; TITLE OF INVENTION: GENES REGULATED BY MYCN ACTIVATION
; FILE REFERENCE: PA-0046 US
; CURRENT APPLICATION NUMBER: US/10/084,817
; CURRENT FILING DATE: 2002-02-25
; PRIOR APPLICATION NUMBER: 60/270,784
; PRIOR FILING DATE: 2001-02-23
; NUMBER OF SEQ ID NOS: 365
; SOFTWARE: PERL Program
; SEQ ID NO 304
; LENGTH: 4826
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030119009A1 238854.23
US-10-084-817-304

Query Match          51.9%; Score 21.8; DB 6; Length 4826;
Best Local Similarity 70.7%; Pred. No. 1e+02;
Matches 29; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAA 41
Db 4091 GCGAAGTTTCACAGAGTCTTCAAAAGACTCTGAAGAAA 4051

RESULT 60
US-09-764-847-1315
; Sequence 1315, Application US/09764847
; Patent No. US20020132767A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009
; CURRENT APPLICATION NUMBER: US/09/764,847
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2003
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1315
; LENGTH: 17849
; TYPE: DNA
; ORGANISM: Homo sapiens
```

```
US-09-764-847-1315

Query Match          51.9%; Score 21.8; DB 3; Length 17849;
Best Local Similarity 70.7%; Pred. No. 1.4e+02;
Matches 29; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 2 CCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAA 42
Db 13871 CCCCACTTCTAAGAGGCTTTATAAAAACTTTCAAGTAA 13911

RESULT 61
US-09-764-891-10139
; Sequence 10139, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10139
; LENGTH: 17849
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-10139

Query Match          51.9%; Score 21.8; DB 3; Length 17849;
Best Local Similarity 70.7%; Pred. No. 1.4e+02;
Matches 29; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 2 CCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAA 42
Db 13871 CCCCACTTCTAAGAGGCTTTATAAAAACTTTCAAGTAA 13911

RESULT 62
US-10-092-154-1315
; Sequence 1315, Application US/10092154
; Publication No. US20030054375A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009C1
; CURRENT APPLICATION NUMBER: US/10/092,154
; CURRENT FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 2003
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1315
; LENGTH: 17849
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-092-154-1315

Query Match          51.9%; Score 21.8; DB 6; Length 17849;
Best Local Similarity 70.7%; Pred. No. 1.4e+02;
Matches 29; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 2 CCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAA 42
Db 13871 CCCCACTTCTAAGAGGCTTTATAAAAACTTTCAAGTAA 13911

RESULT 63
US-09-764-847-1313
; Sequence 1313, Application US/09764847
; Patent No. US20020132767A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
```





Query Match 51.4%; Score 21.6; DB 9; Length 70000;  
Best Local Similarity 75.0%; Pred. No. 2.3e+02;  
Matches 27; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAGCCCTTACAAAGACTATGA 36  
| | | | | | | | | | | | | | | | | | | | | |  
Db 42352 GTCCAGGCTTGTCAGGCCCTTAGAAGGCTAGCAA 42317

## RESULT 68

US-11-097-143-1964  
; Sequence 1964, Application US/11097143  
; Publication No. US20050208558A1  
; GENERAL INFORMATION:  
; APPLICANT: Venter, J. Craig  
; APPLICANT: et al.  
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID  
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE  
; TITLE OF INVENTION: DROSOPHILA GENES.  
; FILE REFERENCE: CL000728  
; CURRENT APPLICATION NUMBER: US/11/097,143  
; CURRENT FILING DATE: 2005-04-04  
; PRIOR APPLICATION NUMBER: 60/157,832  
; PRIOR FILING DATE: 1999-10-05  
; PRIOR APPLICATION NUMBER: 60/160,191  
; PRIOR FILING DATE: 1999-10-19  
; PRIOR APPLICATION NUMBER: 60/161,932  
; PRIOR FILING DATE: 1999-10-28  
; PRIOR APPLICATION NUMBER: 60/164,769  
; PRIOR FILING DATE: 1999-11-12  
; PRIOR APPLICATION NUMBER: 60/173,383  
; PRIOR FILING DATE: 1999-12-28  
; PRIOR APPLICATION NUMBER: 60/175,693  
; PRIOR FILING DATE: 2000-01-12  
; PRIOR APPLICATION NUMBER: 60/184,831  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: 60/191,637  
; PRIOR FILING DATE: 2000-03-23  
; NUMBER OF SEQ ID NOS: 43008  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1964  
; LENGTH: 2356  
; TYPE: DNA  
; ORGANISM: DROSOPHILA

## US-11-097-143-1964

Query Match 51.0%; Score 21.4; DB 13; Length 2356;  
Best Local Similarity 71.8%; Pred. No. 1.3e+02;  
Matches 28; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAGCCCTTACAAAGACTATGA 39  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1212 GCCCAAGAATCGCAGAGCCATTATCAAGAATATCAAGGA 1250

## RESULT 69

US-11-096-568A-26327  
; Sequence 26327, Application US/11096568A  
; Publication No. US20060048249A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; TITLE OF INVENTION: Therby  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096,568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 26327  
; LENGTH: 1227  
; TYPE: DNA  
; ORGANISM: Zea mays subsp. mays  
; FEATURE:

; NAME/KEY: misc feature  
; LOCATION: (1)-(1227)  
; OTHER INFORMATION: Ceres Seq. ID no. 13501189  
US-11-096-568A-26327

Query Match 50.5%; Score 21.2; DB 16; Length 1227;  
Best Local Similarity 69.0%; Pred. No. 1.3e+02;  
Matches 29; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42  
| | | | | | | | | | | | | | | | | | | | | |  
Db 251 GCTTACGCATGGCTGAGCCATAACACACAGATGTTGATGACAAG 292

## RESULT 70

US-10-450-763-28218/c  
; Sequence 28218, Application US/10450763  
; Publication No. US20050196754A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc  
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES  
; FILE REFERENCE: 790CIP3/US  
; CURRENT APPLICATION NUMBER: US/10/450,763  
; CURRENT FILING DATE: 2003-06-11  
; PRIOR APPLICATION NUMBER: PCT/US01/08631  
; PRIOR FILING DATE: 2001-03-30  
; PRIOR APPLICATION NUMBER: 09/540,217  
; PRIOR FILING DATE: 2000-03-31  
; PRIOR APPLICATION NUMBER: 09/649,167  
; PRIOR FILING DATE: 2000-08-23  
; NUMBER OF SEQ ID NOS: 60736  
; SOFTWARE: Custom  
; SEQ ID NO 28218  
; LENGTH: 1446  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SIMILAR  
; LOCATION: (651)-(1361)  
; OTHER INFORMATION: 100% homologous to Homo sapiens dJ185D5.1 (novel  
; OTHER INFORMATION: gene), accession number AL118498, Smith-Waterman Score=1257.  
US-10-450-763-28218

Query Match 50.5%; Score 21.2; DB 10; Length 1446;  
Best Local Similarity 76.5%; Pred. No. 1.4e+02;  
Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAGCCCTTACAAAGACTATG 34  
| | | | | | | | | | | | | | | | | | | | | |  
Db 204 GCCTTCGCTCGCGGAGCCCTCAGGAAGACTATG 171

## RESULT 71

US-10-369-493-36568/c  
; Sequence 36568, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 36568  
; LENGTH: 1555  
; TYPE: DNA

```
; ORGANISM: Aspergillus nidulans
US-10-369-493-36568

Query Match          50.5%; Score 21.2; DB 7; Length 1555;
Best Local Similarity 76.5%; Pred. No. 1.4e+02;
Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 2 CCACGCTTGGCGAGCCCTTACAAAGACTATGA 35
||| ||||| ||||| ||||| |||||
Db 299 CCAACATTGGCGGTACCTTACGCTGACTATGA 266

RESULT 72
US-10-425-115-130787/c
; Sequence 130787, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihwei
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 130787
; LENGTH: 3141
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_50754C.1
US-10-425-115-130787

Query Match          50.5%; Score 21.2; DB 9; Length 3141;
Best Local Similarity 69.0%; Pred. No. 1.7e+02;
Matches 29; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 GCCACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 42
||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1297 GCTTAGCATGGCTGAGCCATACACAGATGTTGATGACAG 1256

RESULT 73
US-09-814-353-3621
; Sequence 3621, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3621
; LENGTH: 527
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 512..519
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-3621

Query Match          50.0%; Score 21; DB 3; Length 527;
Best Local Similarity 73.0%; Pred. No. 1.3e+02;
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 5 ACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAA 41
||| ||||| ||||| ||||| ||||| |||||
Db 458 AAGTTTCACAGAGTCCTTCAAAAGACTTTGAAGAAAA 494

RESULT 75
US-09-814-353-5769
; Sequence 5769, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
```

```
; LENGTH: 527
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 512..519
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-3621

Query Match          50.0%; Score 21; DB 3; Length 527;
Best Local Similarity 73.0%; Pred. No. 1.3e+02;
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 5 ACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAA 41
||| ||||| ||||| ||||| ||||| |||||
Db 458 AAGTTTCACAGAGTCCTTCAAAAGACTTTGAAGAAAA 494

RESULT 74
US-09-814-353-9935
; Sequence 9935, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9935
; LENGTH: 527
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 512..519
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-9935

Query Match          50.0%; Score 21; DB 3; Length 527;
Best Local Similarity 73.0%; Pred. No. 1.3e+02;
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 5 ACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAA 41
||| ||||| ||||| ||||| ||||| |||||
Db 458 AAGTTTCACAGAGTCCTTCAAAAGACTTTGAAGAAAA 494

RESULT 75
US-09-814-353-5769
; Sequence 5769, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
```

```
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5769
; LENGTH: 559
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 514
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-5769

Query Match      50.0%; Score 21; DB 3; Length 559;
Best Local Similarity 73.0%; Pred. No. 1.3e+02;
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy      5  ACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAA 41
      |||||
Db      458 AAGTTTCACAGAGTCCTTCAAAGAGACTCTGAAGAAAA 494

RESULT 76
US-09-814-353-12052
; Sequence 12052, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12052
; LENGTH: 559
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
```

```
; LOCATION: 514
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-12052

Query Match      50.0%; Score 21; DB 3; Length 559;
Best Local Similarity 73.0%; Pred. No. 1.3e+02;
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy      5  ACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAA 41
      |||||
Db      458 AAGTTTCACAGAGTCCTTCAAAGAGACTCTGAAGAAAA 494

RESULT 77
US-09-814-353-4193
; Sequence 4193, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4193
; LENGTH: 579
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 515
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-4193

Query Match      50.0%; Score 21; DB 3; Length 579;
Best Local Similarity 73.0%; Pred. No. 1.4e+02;
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy      5  ACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAA 41
      |||||
Db      459 AAGTTTCACAGAGTCCTTCAAAGAGACTCTGAAGAAAA 495

RESULT 78
US-09-814-353-10499
; Sequence 10499, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
```

```
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; SOFTWARE: FastSeq for Windows Version 4.0
; NUMBER OF SEQ ID NOS: 22037
; SEQ ID NO 10499
; LENGTH: 579
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 515
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-10499
```

```
Query Match 50.0%; Score 21; DB 3; Length 579;
Best Local Similarity 73.0%; Pred. No. 1.4e+02;
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
```

```
Qy 5 ACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAA 41
| | | | | | | | | | | | | | | | | | | | |
Db 459 AAGTTTCACAGAGTCCTTCAAAAGACTTTGAAGAAA 495
```

## RESULT 79

```
US-09-814-353-18436
; Sequence 18436, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18436
; LENGTH: 625
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 606
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-18436
```

```
Query Match 50.0%; Score 21; DB 3; Length 625;
```

```
Best Local Similarity 73.0%; Pred. No. 1.4e+02;
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
```

```
Qy 5 ACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAA 41
| | | | | | | | | | | | | | | | | | | | |
Db 538 AAGTTTCACAGAGTCCTTCAAAAGACTTTGAAGAAA 574
```

## RESULT 80

```
US-09-814-353-18787/c
; Sequence 18787, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18787
; LENGTH: 692
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-814-353-18787
```

```
Query Match 50.0%; Score 21; DB 3; Length 692;
Best Local Similarity 73.0%; Pred. No. 1.4e+02;
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
```

```
Qy 5 ACGCTTGGCGAGCCCTTACAAAGACTATGAAGTAAA 41
| | | | | | | | | | | | | | | | | | | | |
Db 521 AAGTTTCACAGAGTCCTTCAAAAGACTTTGAAGAAA 485
```

## RESULT 81

```
US-09-814-353-20633
; Sequence 20633, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
```

; PRIOR APPLICATION NUMBER: US 60/220,661  
; PRIOR FILING DATE: 2000-07-25  
; PRIOR APPLICATION NUMBER: US 60/257,672  
; PRIOR FILING DATE: 2000-12-21  
; NUMBER OF SEQ ID NOS: 22037  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 20633  
; LENGTH: 1096  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 1, 2, 3, 1096  
; OTHER INFORMATION: n = A,T,C or G  
US-09-814-353-20633

Query Match 50.0%; Score 21; DB 3; Length 1096;  
Best Local Similarity 73.0%; Pred. No. 1.6e+02;  
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 5 ACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAA 41  
Db 437 AAGTTTCACAGAGCTCTTCAAAGAGACTCTGAAGAAAA 473

## RESULT 82

US-09-814-353-20810  
; Sequence 20810, Application US/09814353  
; Publication No. US20030165831A1  
; GENERAL INFORMATION:  
; APPLICANT: Thompson, Pamela  
; APPLICANT: Lillie, James  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR  
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER  
; FILE REFERENCE: MRI-006B

; CURRENT APPLICATION NUMBER: US/09/814,353  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: US 60/191,031  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: US 60/207,124  
; PRIOR FILING DATE: 2000-05-25  
; PRIOR APPLICATION NUMBER: US 60/211,940  
; PRIOR FILING DATE: 2000-06-15  
; PRIOR APPLICATION NUMBER: US 60/216,820  
; PRIOR FILING DATE: 2000-07-07  
; PRIOR APPLICATION NUMBER: US 60/220,661  
; PRIOR FILING DATE: 2000-07-25  
; PRIOR APPLICATION NUMBER: US 60/257,672  
; PRIOR FILING DATE: 2000-12-21  
; NUMBER OF SEQ ID NOS: 22037  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 20810  
; LENGTH: 1386  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 1383, 1384, 1385, 1386  
; OTHER INFORMATION: n = A,T,C or G  
US-09-814-353-20810

Query Match 50.0%; Score 21; DB 3; Length 1386;  
Best Local Similarity 73.0%; Pred. No. 1.7e+02;  
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 5 ACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAA 41  
Db 566 AAGTTTCACAGAGCTCTTCAAAGAGACTCTGAAGAAAA 602

## RESULT 83

US-09-814-353-20790  
; Sequence 20790, Application US/09814353  
; Publication No. US20030165831A1  
; GENERAL INFORMATION:  
; APPLICANT: Thompson, Pamela  
; APPLICANT: Lillie, James  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR  
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER  
; FILE REFERENCE: MRI-006B

; CURRENT APPLICATION NUMBER: US/09/814,353  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: US 60/191,031  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: US 60/207,124  
; PRIOR FILING DATE: 2000-05-25  
; PRIOR APPLICATION NUMBER: US 60/211,940  
; PRIOR FILING DATE: 2000-06-15  
; PRIOR APPLICATION NUMBER: US 60/216,820  
; PRIOR FILING DATE: 2000-07-07  
; PRIOR APPLICATION NUMBER: US 60/220,661  
; PRIOR FILING DATE: 2000-07-25  
; PRIOR APPLICATION NUMBER: US 60/257,672  
; PRIOR FILING DATE: 2000-12-21  
; NUMBER OF SEQ ID NOS: 22037  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 20790  
; LENGTH: 2145  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 1, 2, 2114, 2130, 2138, 2144, 2145  
; OTHER INFORMATION: n = A,T,C or G  
US-09-814-353-20790

Query Match 50.0%; Score 21; DB 3; Length 2145;  
Best Local Similarity 73.0%; Pred. No. 1.8e+02;  
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 5 ACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAA 41  
Db 657 AAGTTTCACAGAGCTCTTCAAAGAGACTCTGAAGAAAA 693

## RESULT 84

US-10-006-285-420/c  
; Sequence 420, Application US/10006285  
; Publication No. US20030165854A1  
; GENERAL INFORMATION:  
; APPLICANT: Mary Jane Cunningham  
; APPLICANT: Matthew R. Kaser  
; TITLE OF INVENTION: MARKER GENES RESPONDING TO TREATMENT WITH TOXINS  
; FILE REFERENCE: PA-0039 US

; CURRENT APPLICATION NUMBER: US/10/006,285  
; CURRENT FILING DATE: 2001-12-05  
; NUMBER OF SEQ ID NOS: 514  
; SOFTWARE: PERL Program  
; SEQ ID NO 420  
; LENGTH: 4605  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. US20030165854A1 238854.22  
US-10-006-285-420

Query Match 50.0%; Score 21; DB 7; Length 4605;  
Best Local Similarity 73.0%; Pred. No. 2.2e+02;  
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 5 ACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAA 41

Db

3959 AAGTTTCACAGAGTCCCTTCAAAGACTCTGAAGAAA 3923

```

RESULT 85
US-10-723-860-5576
; Sequence 5576, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Gineburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5576
; LENGTH: 6346
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (6249)..(6299)
; OTHER INFORMATION: n is a, c, g, or t
; US-10-723-860-5576

```

```

Query Match      50.0%; Score 21; DB 9; Length 6346;
Best Local Similarity 73.0%; Pred.No. 2.4e+02;
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 5 ACGTTGGCGGAGCCCTTACAAAGACTATGAAGTAAA 41
Db 659 AAGTTTTCACAGAGTCCTTCAAAAGACTCTGAAGAAA 695

```

```

RESULT 86
US-10-052-482-214/c
; Sequence 214, Application US/10052482
; Publication No. US20040072264A1
; GENERAL INFORMATION:
; APPLICANT: Encelhard, Eric
; APPLICANT: Morris, David
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: A-71087/RMS/DCF
; CURRENT APPLICATION NUMBER: US/10/052,482
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 241
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 214
; LENGTH: 96589
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (9857)..(9876)
; OTHER INFORMATION: "n" at positions 9857 to 9876 can be any base
US-10-052-482-214

```

```

Query Match      50.0%; Score 21; DB 8; Length 96589;
Best Local Similarity 73.0%; Pred. No. 4.6e+02;
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy:      5  ACGTTTGGCGAGGCCCTTACAAAGACTATATGAAGTAAA 41
          |||||

```

db 45448 ACTCTGTGCCAGCCTGTACAGAAATTATGAAGAAA 45412

```

RESULT 87
US-09-925-065A-148760/c
; Sequence 148760, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108927.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 148760
; LENGTH: 560
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-148760

```

Query Match	49.5%	Score 20.8	DB 4	Length 560
Best Local Similarity	78.1%	Pred. No. 1.6e+02		
Matches	25	Conservative	0	Mismatches 7; Indels 0; Gaps 0
QY	2	CCCAACGTTGGCGGAGCCCTTACAAAGACTAT	33	
Db	35	CCCAACCTTGGCAGAGCCATCAAGTACTAT	4	

```

RESULT 88
US-09-925-065A-148761/c
; Sequence 148761, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FAST-SEQ for Windows Version 4.0
; SEQ ID NO 148761
; LENGTH: 560
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-148761

```

Query Match	49.5%	Score 20.8;	DB 4;	Length 560;
Best Local Similarity	78.1%	Pred. No. 1.6e+02;		
Matches 25; Conservative		0; Mismatches 7;	Indels 0;	Gaps 0;

Qy 2 CCCACGCTTGGCGAGCCCTTACAAAGACTAT 33  
||||| ||||| ||||| ||||| ||||| |||||  
Db 35 CCCAACCTTGGCAGAGCCCATCACAAGTACTAT 4

## RESULT 89

US-09-925-065A-148760/c  
; Sequence 148760, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925.065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 148760  
; LENGTH: 560  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-148760

Query Match 49.5%; Score 20.8; DB 5; Length 560;  
Best Local Similarity 78.1%; Pred. No. 1.6e+02;  
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 2 CCCACGCTTGGCGAGCCCTTACAAAGACTAT 33  
||||| ||||| ||||| ||||| ||||| |||||  
Db 35 CCCAACCTTGGCAGAGCCCATCACAAGTACTAT 4

## RESULT 90

US-09-925-065A-148761/c  
; Sequence 148761, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925.065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 148761  
; LENGTH: 560  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-148761

Query Match 49.5%; Score 20.8; DB 5; Length 560;  
Best Local Similarity 78.1%; Pred. No. 1.6e+02;  
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 2 CCCACGCTTGGCGAGCCCTTACAAAGACTAT 33  
||||| ||||| ||||| ||||| ||||| |||||  
Db 35 CCCAACCTTGGCAGAGCCCATCACAAGTACTAT 4

## RESULT 91

US-10-301-480-242891/c  
; Sequence 242891, Application US/10301480  
; Publication No. US20060057564A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms  
; FILE REFERENCE: 108827.137  
; CURRENT APPLICATION NUMBER: US/10/301.480  
; CURRENT FILING DATE: 2002-11-21  
; PRIOR APPLICATION NUMBER: US 10/215,598  
; PRIOR FILING DATE: 2002-08-09  
; PRIOR APPLICATION NUMBER: US 60/311,695  
; PRIOR FILING DATE: 2001-08-10  
; NUMBER OF SEQ ID NOS: 1226818  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 242891  
; LENGTH: 569  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-10-301-480-242891

Query Match 49.5%; Score 20.8; DB 12; Length 569;  
Best Local Similarity 78.1%; Pred. No. 1.7e+02;  
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 2 CCCACGCTTGGCGAGCCCTTACAAAGACTAT 33  
||||| ||||| ||||| ||||| ||||| |||||  
Db 44 CCCAACCTTGGCAGAGCCCATCACAAGTACTAT 13

## RESULT 92

US-10-301-480-242892/c  
; Sequence 242892, Application US/10301480  
; Publication No. US20060057564A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms  
; FILE REFERENCE: 108827.137  
; CURRENT APPLICATION NUMBER: US/10/301.480  
; CURRENT FILING DATE: 2002-11-21  
; PRIOR APPLICATION NUMBER: US 10/215,598  
; PRIOR FILING DATE: 2002-08-09  
; PRIOR APPLICATION NUMBER: US 60/311,695  
; PRIOR FILING DATE: 2001-08-10  
; NUMBER OF SEQ ID NOS: 1226818  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 242892  
; LENGTH: 569  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-10-301-480-242892

Query Match 49.5%; Score 20.8; DB 12; Length 569;  
Best Local Similarity 78.1%; Pred. No. 1.7e+02;  
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 2 CCCACGCTTGGCGAGCCCTTACAAAGACTAT 33  
||||| ||||| ||||| ||||| ||||| |||||  
Db 44 CCCAACCTTGGCAGAGCCCATCACAAGTACTAT 13

```
RESULT 93
US-10-301-480-856300/c
; Sequence 856300, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms
; FILE OF INVENTION: in the Human Genome
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 856300
; LENGTH: 569
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-856300
Query Match 49.5%; Score 20.8; DB 12; Length 569;
Best Local Similarity 78.1%; Pred. No. 1.7e+02;
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 CCCACGCTTGGCCGAGCCCTTACAAAGACTAT 33
Db 44 CCCAACCTTGGCAGAGCCATCACAAGTACTAT 13

RESULT 94
US-10-301-480-856301/c
; Sequence 856301, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms
; FILE OF INVENTION: in the Human Genome
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 856301
; LENGTH: 569
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-856301
Query Match 49.5%; Score 20.8; DB 12; Length 569;
Best Local Similarity 78.1%; Pred. No. 1.7e+02;
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 CCCACGCTTGGCCGAGCCCTTACAAAGACTAT 33
Db 44 CCCAACCTTGGCAGAGCCATCACAAGTACTAT 13

RESULT 95
US-09-925-065A-716598/c
; Sequence 716598, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 716598
; LENGTH: 1465
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-716598
Query Match 49.5%; Score 20.8; DB 4; Length 1465;
Best Local Similarity 70.0%; Pred. No. 2.1e+02;
Matches 28; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 3 CCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42
Db 1184 CCGTGCTTGGCCGAGATCCTTAATGACTAGTAGTATGGAGAG 1145

RESULT 96
US-09-925-065A-716599/c
; Sequence 716599, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 716599
; LENGTH: 1465
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-716599
Query Match 49.5%; Score 20.8; DB 4; Length 1465;
Best Local Similarity 70.0%; Pred. No. 2.1e+02;
Matches 28; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 3 CCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 42
Db 1184 CCGTGCTTGGCCGAGATCCTTAATGACTAGTAGTATGGAGAG 1145

RESULT 97
US-09-925-065A-716600/c
; Sequence 716600, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
```





Search completed: June 10, 2006, 20:50:45  
Job time : 1636.6 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2006, 15:18:44 ; Search time 463.533 Seconds  
(without alignments)  
2345.260 Million cell updates/sec

Title: US-10-600-816-31

Perfect score: 17

Sequence: 1 gccacgctggccgag 17

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 6366136 seqs, 31973710525 residues

Total number of hits satisfying chosen parameters: 12732272

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 1000 summaries

Database :

GenEmbl.\*

1: gb\_env.\*

2: gb\_pat.\*

3: gb\_ph.\*

4: gb\_pl.\*

5: gb\_pr.\*

6: gb\_ro.\*

7: gb\_sts.\*

8: gb\_sv.\*

9: gb\_un.\*

10: gb\_vi.\*

11: gb\_ov.\*

12: gb\_htg.\*

13: gb\_in.\*

14: gb\_om.\*

15: gb\_ba.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Length	ID	Description
1	17	100.0	603	BD233463 Human pro
2	17	100.0	1212	BD209699 Compositi
3	17	100.0	1212	AX341505 Sequence
4	17	100.0	1619	AX078375 Sequence
5	17	100.0	1718	BD233473 Human pro
6	17	100.0	1826	AK172760 Homo sapi
7	17	100.0	2290	C0723177 Sequence
8	17	100.0	2296	BC003665 Homo sapi
9	17	100.0	2297	AF506289 Homo sapi
10	17	100.0	2302	AX930411 Sequence
11	17	100.0	2302	AF095448 Homo sapi
12	17	100.0	2446	BD156680 Primer fo
13	17	100.0	2446	AX877483 Sequence
14	17	100.0	2446	AK001761 Homo sapi
15	17	100.0	2456	C0981495 Sequence
16	17	100.0	2456	DD210040 Methods o
17	17	100.0	2456	AX549168 Sequence
18	17	100.0	3057	AK122672 Homo sapi

19	17	100.0	3371	2	AX188348	Sequence
20	17	100.0	6730	2	CQ894732	Sequence
c 21	17	100.0	161577	5	AC007688	Homo sapi
c 22	17	100.0	213601	12	AC165037	Bos tauru
23	17	100.0	284106	12	AC166087	Bos tauru
c 24	17	100.0	295050	15	SC039104	Streptomy
25	16	94.1	110000	15	CP000124_29	Continuation (30 o
26	16	94.1	110000	15	BX571965_26	Continuation (27 o
27	16	94.1	110000	15	CP000010_16	Continuation (17 o
c 28	16	94.1	174298	12	AC167299	Oryctolag
c 29	15.4	90.6	117	15	MAU564400	Mycobacte
30	15.4	90.6	385	5	AM181332	Homo sapi
c 31	15.4	90.6	586	1	AY535046	Unculture
c 32	15.4	90.6	738	10	DQ288979	Bovine le
c 33	15.4	90.6	1802	1	AY381191	Unculture
c 34	15.4	90.6	2306	2	AX416413	Sequence
35	15.4	90.6	2550	5	HSAS64991	Homo sapi
36	15.4	90.6	2550	5	HSAS64992	Homo sapi
37	15.4	90.6	2550	5	HSAS64993	Homo sapi
38	15.4	90.6	2623	5	BC018939	Homo sapi
39	15.4	90.6	2623	5	BC034926	Homo sapi
40	15.4	90.6	3252	4	AF420019	Aspergill
41	15.4	90.6	3325	2	CQ727552	Sequence
42	15.4	90.6	3602	5	AF542080	Homo sapi
43	15.4	90.6	4183	2	AR157945	Sequence
44	15.4	90.6	4183	2	BD103585	Human nuc
45	15.4	90.6	4232	2	CQ850573	Sequence
46	15.4	90.6	4232	5	AK127735	Homo sapi
47	15.4	90.6	4641	5	AB209580	Homo sapi
48	15.4	90.6	5425	5	AY375870	Homo sapi
49	15.4	90.6	5599	5	AY375871	Homo sapi
50	15.4	90.6	5600	5	AY375872	Homo sapi
51	15.4	90.6	5600	5	AY375873	Homo sapi
c 52	15.4	90.6	8714	10	BLVCG	Bovine leuk
c 53	15.4	90.6	10694	15	AE004474	Pseudomon
54	15.4	90.6	16651	4	AB168089	Chlamydom
55	15.4	90.6	59316	6	BX000454	Mouse DNA
56	15.4	90.6	6661	12	AC165982	Bos tauru
57	15.4	90.6	99966	5	AY663411	Homo sapi
58	15.4	90.6	100761	5	AY663415	Homo sapi
c 59	15.4	90.6	103216	5	HS039318	Human DNA
60	15.4	90.6	103395	5	AY663406	Homo sapi
61	15.4	90.6	104996	5	AY663395	Homo sapi
c 62	15.4	90.6	110000	15	AE001439_02	Continuation (3 of
c 63	15.4	90.6	110000	15	AE001439_03	Continuation (4 of
c 64	15.4	90.6	110000	15	CP000240_27	Continuation (28 o
c 65	15.4	90.6	110000	15	CP000240_28	Continuation (29 o
c 66	15.4	90.6	110000	15	CP000248_28	Continuation (29 o
c 67	15.4	90.6	110000	15	AE014295_08	Continuation (9 of
c 68	15.4	90.6	110000	15	AE017262_13	Continuation (14 o
c 69	15.4	90.6	110000	15	BA000032_12	Continuation (13 o
c 70	15.4	90.6	113733	5	AP003159	Homo sapi
c 71	15.4	90.6	147557	5	AL662789	Human DNA
c 72	15.4	90.6	150372	12	AC150103	Gallus ga
73	15.4	90.6	150375	12	AC169542	Bos tauru
74	15.4	90.6	153108	5	AL355273	Human DNA
c 75	15.4	90.6	160714	12	AC150149	Gallus ga
76	15.4	90.6	185122	2	AC011448	Homo sapi
77	15.4	90.6	185122	12	AC146997	Arbacta p
c 78	15.4	90.6	172995	6	AC130666	Mus muscu
c 79	15.4	90.6	179959	5	AL160395	Human DNA
c 80	15.4	90.6	183887	5	AC017033	Homo sapi
81	15.4	90.6	187027	12	AC019078	Homo sapi
c 82	15.4	90.6	193015	12	AC150074	Gallus ga
83	15.4	90.6	195269	2	AX417035	Sequence
c 84	15.4	90.6	196542	5	AC002365	Homo sapi
c 85	15.4	90.6	197456	12	AC005054	Homo sapi
c 86	15.4	90.6	198375	6	AC120873	Mus muscu
c 87	15.4	90.6	200618	5	AC009061	Homo sapi
88	15.4	90.6	205915	6	AC119828	Mus muscu
c 89	15.4	90.6	209262	12	AC021091	Homo sapi
90	15.4	90.6	214798	6	AC091522	Mus muscu
91	15.4	90.6	216540	12	AC134138	Rattus no





C 384	14	82.4	183113	5	HS243E7	AL0232323 Human DNA	457	13.8	81.2	407	10	DQ222885	Human her
C 385	14	82.4	183402	12	AC126231	AC126231 Bos tauru	458	13.8	81.2	407	10	DQ222886	Human her
C 386	14	82.4	188030	12	AC015456	AC015456 Homo sapi	459	13.8	81.2	407	10	DQ223889	Human her
C 387	14	82.4	189174	5	AC092687	AC092687 Homo sapi	460	13.8	81.2	407	10	DQ223890	Human her
C 388	14	82.4	189360	12	AC146696	AC146696 Pan trogl	461	13.8	81.2	407	10	DQ223891	Human her
C 389	14	82.4	189381	12	AC126920	AC126920 Bos tauru	462	13.8	81.2	407	10	DQ223892	Human her
C 390	14	82.4	190200	6	AC114613	AC114613 Mus muscu	463	13.8	81.2	407	10	DQ223893	Human her
C 391	14	82.4	192653	5	AC124781	AC124781 Homo sapi	464	13.8	81.2	407	10	DQ224079	Human her
C 392	14	82.4	195275	6	AC102255	AC102255 Mus muscu	465	13.8	81.2	407	10	DQ224080	Human her
C 393	14	82.4	197719	6	AC168212	AC168212 Mus muscu	466	13.8	81.2	407	10	DQ224081	Human her
C 394	14	82.4	199454	12	AC117180	AC117180 Homo sapi	467	13.8	81.2	407	10	DQ231011	Human her
C 395	14	82.4	198673	6	AL669872	AL669872 Mouse DNA	468	13.8	81.2	407	10	DQ231012	Human her
C 396	14	82.4	199928	12	AC120941	AC120941 Rattus no	469	13.8	81.2	407	10	DQ231013	Human her
C 397	14	82.4	201269	12	AP001849	AP001849 Homo sapi	470	13.8	81.2	407	10	DQ231014	Human her
C 398	14	82.4	201629	12	AL158047	AL158047 Human DNA	471	13.8	81.2	407	10	DQ231015	Human her
C 399	14	82.4	201673	6	AL591544	AL591544 Mouse DNA	472	13.8	81.2	407	10	DQ231016	Human her
C 400	14	82.4	202402	12	AC012517	AC012517 Homo sapi	473	13.8	81.2	407	10	DQ231017	Human her
C 401	14	82.4	204963	12	AC163566	AC163566 Bos tauru	474	13.8	81.2	407	10	DQ231018	Human her
C 402	14	82.4	209121	6	AC127589	AC127589 Mus muscu	475	13.8	81.2	407	10	DQ231019	Human her
C 403	14	82.4	209775	12	AC027219	AC027219 Homo sapi	476	13.8	81.2	407	10	DQ231020	Human her
C 404	14	82.4	211209	12	CT485990	CT485990 Drosophil	477	13.8	81.2	407	10	DQ231021	Human her
C 405	14	82.4	211531	5	AC006995	AC006995 Homo sapi	478	13.8	81.2	407	10	DQ231022	Human her
C 406	14	82.4	213925	12	AC150641	AC150641 Bos tauru	479	13.8	81.2	407	10	DQ231023	Human her
C 407	14	82.4	216211	12	AC091315	AC091315 Mus muscu	480	13.8	81.2	407	10	DQ231024	Human her
C 408	14	82.4	217979	12	AC161613	AC161613 Pan trogl	481	13.8	81.2	407	10	DQ231025	Human her
C 409	14	82.4	220289	6	AC108908	AC108908 Mus muscu	482	13.8	81.2	408	10	DQ223894	Human her
C 410	14	82.4	220932	5	BS000125	BS000125 Pan trogl	483	13.8	81.2	417	2	C0604581	Sequence
C 411	14	82.4	224337	6	AC163434	AC163434 Mus muscu	484	13.8	81.2	420	2	BD262970	DNA seq
C 412	14	82.4	228292	6	AL772199	AL772199 Mouse DNA	485	13.8	81.2	420	2	AX024262	Sequence
C 413	14	82.4	229508	5	AC146694	AC146694 Pan trogl	486	13.8	81.2	420	2	AX024262	Sequence
C 414	14	82.4	229898	12	AC161564	AC161564 Bos tauru	487	13.8	81.2	420	2	HSC59C06	Sequence
C 415	14	82.4	230552	5	AC005098	AC005098 Homo sapi	488	13.8	81.2	420	2	AF1133135	Sordaria
C 416	14	82.4	232945	12	AC118407	AC118407 Rattus no	489	13.8	81.2	474	4	SF1133136	Sordaria
C 417	14	82.4	235707	12	AC169692	AC169692 Bos tauru	490	13.8	81.2	501	13	AF060851	Haliotis
C 418	14	82.4	239307	12	AC129402	AC129402 Rattus no	491	13.8	81.2	515	4	AY558742	Balaemia
C 419	14	82.4	246322	12	AC133966	AC133966 Homo sapi	492	13.8	81.2	521	10	AF134734	Hepatit
C 420	14	82.4	246322	12	AC133966	AC133966 Homo sapi	493	13.8	81.2	521	10	AF134734	Hepatit
C 421	14	82.4	247376	12	AC156127	AC156127 Bos tauru	494	13.8	81.2	521	10	AF134734	Hepatit
C 422	14	82.4	250425	12	AC125667	AC125667 Rattus no	495	13.8	81.2	521	10	AF134734	Hepatit
C 423	14	82.4	268050	15	AL627266	AL627266 Salmonell	496	13.8	81.2	521	10	AF134734	Hepatit
C 424	14	82.4	275197	5	AC004166	AC004166 Homo sapi	497	13.8	81.2	521	10	AF134734	Hepatit
C 425	14	82.4	281138	12	AC159058	AC159058 Bos tauru	498	13.8	81.2	521	10	AF134734	Hepatit
C 426	14	82.4	285354	12	AC152597	AC152597 Bos tauru	499	13.8	81.2	521	10	AF134734	Hepatit
C 427	14	82.4	292748	12	AC106574	AC106574 Rattus no	500	13.8	81.2	521	10	AF134734	Hepatit
C 428	14	82.4	300350	15	BX294145	BX294145 Pirellula	501	13.8	81.2	521	10	AF134734	Hepatit
C 429	14	82.4	302521	5	AP000014	AP000014 Homo sapi	502	13.8	81.2	521	10	AF134734	Hepatit
C 430	14	82.4	340000	5	AP001730	AP001730 Homo sapi	503	13.8	81.2	521	10	AF134734	Hepatit
C 431	14	82.4	349980	2	AX492782	AX492782 Sequence	504	13.8	81.2	521	10	AF134734	Hepatit
C 432	14	82.4	349980	2	AX492786	AX492786 Sequence	505	13.8	81.2	521	10	AF134734	Hepatit
C 433	14	82.4	349980	2	AX553949	AX553949 Sequence	506	13.8	81.2	521	10	AF134734	Hepatit
C 434	14	82.4	349980	2	AX553953	AX553953 Sequence	507	13.8	81.2	521	10	AF134734	Hepatit
C 435	13.8	81.2	20	2	AR568092	AR568092 Sequence	508	13.8	81.2	521	10	AF134734	Hepatit
C 436	13.8	81.2	47	2	AR290691	AR290691 Sequence	509	13.8	81.2	521	10	AF134734	Hepatit
C 437	13.8	81.2	117	15	NPE564413	AJ564413 Mycobacte	510	13.8	81.2	521	10	AF134734	Hepatit
C 438	13.8	81.2	117	15	MXE564419	AJ564419 Mycobacte	511	13.8	81.2	521	10	AF134734	Hepatit
C 439	13.8	81.2	123	5	AF439454	AF439454 Homo sapi	512	13.8	81.2	521	10	AF134734	Hepatit
C 440	13.8	81.2	201	7	BV171711	BV171711 sqm45306	513	13.8	81.2	521	10	AF134734	Hepatit
C 441	13.8	81.2	206	7	CR382032	CR382032 Arabidops	514	13.8	81.2	521	10	AF134734	Hepatit
C 442	13.8	81.2	209	6	WNHIF3A06	AF079145 Mus muscu	515	13.8	81.2	521	10	AF134734	Hepatit
C 443	13.8	81.2	219	10	AF134739	AF134739 Hepatitis	516	13.8	81.2	521	10	AF134734	Hepatit
C 444	13.8	81.2	228	2	AC131160	AC131160 Sequence	517	13.8	81.2	521	10	AF134734	Hepatit
C 445	13.8	81.2	231	5	HS13F5R	Z54608 H.sapiens C	518	13.8	81.2	521	10	AF134734	Hepatit
C 446	13.8	81.2	240	10	HS1LRN1A	K03352 Hsv1 (KOS) ,	519	13.8	81.2	521	10	AF134734	Hepatit
C 447	13.8	81.2	247	8	SYNLRBHC4	M22006 Synthetic h	520	13.8	81.2	521	10	AF134734	Hepatit
C 448	13.8	81.2	273	7	BY0000760	BY0000760 S210P6198	521	13.8	81.2	521	10	AF134734	Hepatit
C 449	13.8	81.2	283	2	AR247212	AR247212 Sequence	522	13.8	81.2	521	10	AF134734	Hepatit
C 450	13.8	81.2	293	2	AR249750	AR249750 Sequence	523	13.8	81.2	521	10	AF134734	Hepatit
C 451	13.8	81.2	320	13	AY465871	AY465871 Clinostat	524	13.8	81.2	521	10	AF134734	Hepatit
C 452	13.8	81.2	350	2	CQ458356	CQ458356 Sequence	525	13.8	81.2	521	10	AF134734	Hepatit
C 453	13.8	81.2	363	2	AX308452	AX308452 Sequence	526	13.8	81.2	521	10	AF134734	Hepatit
C 454	13.8	81.2	375	15	AV726000	AV726000 Pseudomon	527	13.8	81.2	521	10	AF134734	Hepatit
C 455	13.8	81.2	385	2	AR149028	AR149028 Sequence	528	13.8	81.2	521	10	AF134734	Hepatit
C 456	13.8	81.2	385	2	AR340400	AR340400 Sequence	529	13.8	81.2	521	10	AF134734	Hepatit



c 676	13.8	81.2	2165	6	BC090335	BC090335 Rattus no	c 749	13.8	81.2	4317	5	BC040228	BC040228 Homo sapi
c 677	13.8	81.2	2169	2	AR447719	AR447719 Sequence	c 750	13.8	81.2	4319	2	AX405634	AX405634 Sequence
678	13.8	81.2	2174	2	AR225609	AR225609 Sequence	752	13.8	81.2	4364	8	AY497507	AY497507 Cloning v
679	13.8	81.2	2174	2	AR692065	AR692065 Sequence	752	13.8	81.2	4381	2	DD165226	DD165226 Nucleic A
680	13.8	81.2	2174	2	AX365106	AX365106 Sequence	753	13.8	81.2	4399	5	DQ173578	DQ173578 Pan trogl
681	13.8	81.2	2184	2	AR623715	AR623715 Sequence	754	13.8	81.2	4455	11	XLU14164	XLU14164 Xenopus lae
c 682	13.8	81.2	2194	6	BC098750	BC098750 Rattus no	755	13.8	81.2	4561	8	AY745747	AY745747 PGeneClip
c 683	13.8	81.2	2205	6	BC012270	BC012270 Mus muscu	756	13.8	81.2	4563	5	BC041767	BC041767 Homo sapi
c 684	13.8	81.2	2207	6	AF060194	AF060194 Mus muscu	c 757	13.8	81.2	4579	15	PSOPRC	D28119 Pseudomonas
c 685	13.8	81.2	2318	15	AF291753	AF291753 Pseudomon	758	13.8	81.2	4611	15	RSHISAFEG	U27829 R.sphaeroid
c 686	13.8	81.2	2328	15	PSEALG76	M37181 P.aeruginos	c 759	13.8	81.2	4617	15	PAU27829	U27829 Pseudomonas
687	13.8	81.2	2351	2	CS113293	CS113293 Sequence	c 760	13.8	81.2	4625	5	AY605898	AY605898 Homo sapi
688	13.8	81.2	2357	10	AY363172	AY363172 Pseudorab	761	13.8	81.2	4635	8	AF346623	AF346623 RAGE vect
c 689	13.8	81.2	2417	2	CQ604580	CQ604580 Sequence	762	13.8	81.2	4783	2	CQ850207	CQ850207 Sequence
c 690	13.8	81.2	2433	6	AB041557	AB041557 Mus muscu	763	13.8	81.2	4783	5	AK128774	AK128774 Homo sapi
c 691	13.8	81.2	2434	6	AB041559	AB041559 Mus muscu	764	13.8	81.2	4852	2	CQ584930	CQ584930 Sequence
692	13.8	81.2	2477	14	BC112800	BC112800 Bos Tauru	765	13.8	81.2	4951	8	ASAJ154	AY000154 Artificia
693	13.8	81.2	2501	2	CQ778570	CQ778570 Sequence	766	13.8	81.2	5043	2	CS165130	CS165130 Sequence
c 694	13.8	81.2	2510	10	SHU12XA	M95285 Suid herpes	767	13.8	81.2	5049	2	CQ598359	CQ598359 Sequence
c 695	13.8	81.2	2557	4	AY091763	AY091763 Arabidops	768	13.8	81.2	5049	2	CQ847818	CQ847818 Sequence
696	13.8	81.2	2573	2	CQ843419	CQ843419 Sequence	769	13.8	81.2	5101	2	CS165129	CS165129 Sequence
697	13.8	81.2	2573	5	AK126380	AK126380 Homo sapi	770	13.8	81.2	5139	15	AY729020	AY729020 Rhizobium
698	13.8	81.2	2585	4	AK106303	AK106303 Oryza sat	771	13.8	81.2	5177	6	AF071086	AF071086 Mus muscu
699	13.8	81.2	2626	6	AB041588	AB041588 Mus muscu	c 772	13.8	81.2	5204	2	CQ818774	CQ818774 Sequence
700	13.8	81.2	2634	5	HSAS64981	AY564981 Homo sapi	c 773	13.8	81.2	5235	11	AB081299	AB081299 Sililago j
701	13.8	81.2	2787	2	CQ584931	CQ584931 Sequence	774	13.8	81.2	5325	8	ASPRESP	ZV5185 Artificial
702	13.8	81.2	2893	13	AY245701	AY245701 Clinostom	775	13.8	81.2	5351	8	AY024358	AY024358 Integrati
703	13.8	81.2	2893	13	BT023853	BT023853 Drosophil	776	13.8	81.2	5382	2	AR215117	AR215117 Sequence
704	13.8	81.2	2916	15	PME012480	AY012480 Pseudomon	777	13.8	81.2	5382	2	AR302359	AR302359 Sequence
c 705	13.8	81.2	2925	5	HUMCA1IA01	M77176 H.sapiens c	778	13.8	81.2	5382	2	AR373231	AR373231 Sequence
706	13.8	81.2	2974	13	AY242851	AY242851 Bolbophor	779	13.8	81.2	5382	2	AR401617	AR401617 Sequence
707	13.8	81.2	3027	13	AY122082	AY122082 Drosophil	780	13.8	81.2	5382	2	AR442950	AR442950 Sequence
708	13.8	81.2	3110	6	BC013092	BC013092 Mus muscu	781	13.8	81.2	5382	2	AR568592	AR568592 Sequence
709	13.8	81.2	3133	2	I15000	I15000 Sequence 1	782	13.8	81.2	5382	2	AR658613	AR658613 Sequence
710	13.8	81.2	3135	11	GGCRK6A	Z19110 G.gallus Ce	c 783	13.8	81.2	5382	15	AB003906	AB003906 Hydrogeno
711	13.8	81.2	3179	2	CQ849853	CQ849853 Sequence	784	13.8	81.2	5466	5	AB209638	AB209638 Homo sapi
712	13.8	81.2	3179	5	AK128915	AK128915 Homo sapi	c 785	13.8	81.2	5618	15	AY622309	AY622309 Pectobact
713	13.8	81.2	3216	6	BC085251	BC085251 Mus muscu	786	13.8	81.2	5630	5	AY375845	AY375845 Homo sapi
714	13.8	81.2	3297	8	AF090453	AF090453 Cloning v	787	13.8	81.2	5648	5	AY375847	AY375847 Homo sapi
715	13.8	81.2	3357	2	CQ721036	CQ721036 Sequence	788	13.8	81.2	5650	8	DQ172901	DQ172901 Transfect
716	13.8	81.2	3360	2	CQ591030	CQ591030 Sequence	789	13.8	81.2	5656	8	CR847878	CR847878 Condition
c 717	13.8	81.2	3387	11	AB231587	AB231587 Danio rer	790	13.8	81.2	5664	5	AY375851	AY375851 Homo sapi
c 718	13.8	81.2	3438	11	AF508797	AF508797 Gallus ga	791	13.8	81.2	5668	5	AY375852	AY375852 Homo sapi
c 719	13.8	81.2	3453	2	AR623312	AR623312 Sequence	792	13.8	81.2	5693	5	AY375854	AY375854 Homo sapi
720	13.8	81.2	3500	5	AF442769	AF442769 Homo sapi	793	13.8	81.2	5697	5	AY375848	AY375848 Homo sapi
721	13.8	81.2	3544	4	AB020748	AB020748 Arabidops	794	13.8	81.2	5699	5	AY375849	AY375849 Homo sapi
722	13.8	81.2	3558	5	HUMDRBU	M15809 Human MHC c	795	13.8	81.2	5703	5	AY375846	AY375846 Homo sapi
723	13.8	81.2	3575	5	HSHEFH4	X94553 H.sapiens H	796	13.8	81.2	5706	8	AF529183	AF529183 Episomal
c 724	13.8	81.2	3604	6	AF348523	AF348523 Rattus no	797	13.8	81.2	5707	5	AY375853	AY375853 Homo sapi
725	13.8	81.2	3636	2	AX714273	AX714273 Sequence	798	13.8	81.2	5711	5	AY375850	AY375850 Homo sapi
c 726	13.8	81.2	3636	5	AK056495	AK056495 Homo sapi	c 799	13.8	81.2	5936	8	AY613996	AY613996 Cloning v
c 727	13.8	81.2	3657	6	AY749168	AY749168 Mus muscu	c 800	13.8	81.2	6066	2	CQ591029	CQ591029 Sequence
728	13.8	81.2	3668	15	RTNODG	X03721 Rhizobium t	c 801	13.8	81.2	6076	8	AB086386	AB086386 Retrovira
c 729	13.8	81.2	3674	5	AF198358	AF198358 Homo sapi	802	13.8	81.2	6113	15	AF393183	AF393183 Clavibact
c 730	13.8	81.2	3729	2	CQ850157	CQ850157 Sequence	803	13.8	81.2	6233	2	CS165135	CS165135 Sequence
c 731	13.8	81.2	3729	5	AK127261	AK127261 Homo sapi	804	13.8	81.2	6279	2	AX207276	AX207276 Sequence
c 732	13.8	81.2	3747	6	AF013131	AF013131 Mus muscu	805	13.8	81.2	6289	5	AB018339	AB018339 Homo sapi
c 733	13.8	81.2	3752	11	FRD1	X80174 F.rubripes	c 806	13.8	81.2	6381	2	BD171160	BD171160 Novel gen
734	13.8	81.2	3766	13	BT011380	BT011380 Drosophil	c 807	13.8	81.2	6381	2	BD183438	BD183438 Novel gen
c 735	13.8	81.2	3797	10	PVULSGENE	X87247 Pseudorabie	c 808	13.8	81.2	6381	5	AB051439	AB051439 Homo sapi
c 736	13.8	81.2	3814	2	AR531687	AR531687 Sequence	c 809	13.8	81.2	6381	5	AB051439	AB051439 Homo sapi
c 737	13.8	81.2	3956	2	AX801721	AX801721 Sequence	c 810	13.8	81.2	6438	15	MKMDMTD	Y10251 M.kandleri
c 738	13.8	81.2	3956	6	AF074266	AF074266 Mus muscu	c 811	13.8	81.2	6515	2	CS175841	CS175841 Sequence
c 739	13.8	81.2	4040	8	AY613994	AY613994 Cloning v	812	13.8	81.2	6561	2	CS085795	CS085795 Sequence
c 740	13.8	81.2	4071	6	AB011439	AB011439 Rattus no	813	13.8	81.2	6630	13	AY197345	AY197345 Schistosom
c 741	13.8	81.2	4133	10	AB218901	AB218901 Chimpanze	814	13.8	81.2	6637	13	AY197343	AY197343 Schistosom
742	13.8	81.2	4173	8	AY508732	AY508732 Cloning v	815	13.8	81.2	6657	13	AY197344	AY197344 Schistosom
743	13.8	81.2	4192	5	BC043183	BC043183 Homo sapi	816	13.8	81.2	6762	8	AF334709	AF334709 Envelope
c 744	13.8	81.2	4214	5	AB037800	AB037800 Homo sapi	817	13.8	81.2	7122	2	CS175842	CS175842 Sequence
745	13.8	81.2	4257	2	CQ918571	CQ918571 Sequence	c 818	13.8	81.2	7180	2	AX468870	AX468870 Sequence
746	13.8	81.2	4257	8	CVU07648	U07648 Cloning vec	819	13.8	81.2	7201	2	CQ889769	CQ889769 Sequence
747	13.8	81.2	4282	5	DQ173577	DQ173577 Homo sapi	c 820	13.8	81.2	7201	2	CQ973535	CQ973535 Sequence
c 748	13.8	81.2	4307	11	BC079756	BC079756 Xenopus 1	c 821	13.8	81.2	7201	2	AX800664	AX800664 Sequence



C 822	13.8	81.2	7231	2	AX806466	AX806466 Sequence	C 895	13.8	81.2	10637	5	AY184206	AY184206 Homo sapi
823	13.8	81.2	7291	8	ASPSVPZ1	Y07573 pSPa211 pl	C 896	13.8	81.2	10654	8	CR293541	CR293541 3'HPT (M
824	13.8	81.2	7294	8	ASPSVPX1	Y07631 pSPaX1 pla	C 897	13.8	81.2	10742	2	CS036440	CS036440 Sequence
C 825	13.8	81.2	7297	2	AX806467	AX806467 Sequence	C 898	13.8	81.2	10742	2	CS045392	CS045392 Sequence
826	13.8	81.2	7353	4	CNS09540	BX072545 Oryza sat	C 899	13.8	81.2	10742	5	AY061755	AY061755 Homo sapi
C 827	13.8	81.2	7429	2	AX839727	AX839727 Sequence	C 900	13.8	81.2	10774	2	CQ789660	CQ789660 Sequence
828	13.8	81.2	7438	2	AX816382	AX816382 Sequence	C 901	13.8	81.2	10805	12	AC017707	AC017707 Drosophila
C 829	13.8	81.2	7456	2	AX806465	AX806465 Sequence	C 902	13.8	81.2	10868	8	DQ322637	DQ322637 VEEV repl
830	13.8	81.2	7504	5	AB040943	AB040943 Homo sapi	C 903	13.8	81.2	10921	2	CQ789658	CQ789658 Sequence
831	13.8	81.2	7573	2	AX114856	AX114856 Sequence	C 904	13.8	81.2	10928	8	DQ322636	DQ322636 VEEV repl
C 832	13.8	81.2	7573	2	AX114870	AX114870 Sequence	C 905	13.8	81.2	10928	8	DQ322639	DQ322639 VEEV repl
C 833	13.8	81.2	7800	10	HEHSV1EX	X03839 HSV-1 (stra	C 906	13.8	81.2	10928	8	DQ322641	DQ322641 VEEV repl
834	13.8	81.2	8050	15	AE006033	AE006033 Caulobact	C 907	13.8	81.2	10959	15	AE004953	AE004953 Pseudomon
C 835	13.8	81.2	8157	12	AL831743	AL831743 Mus muscu	C 908	13.8	81.2	10961	2	CQ789656	CQ789656 Sequence
836	13.8	81.2	8298	2	A70359	A70359 Sequence 1	C 909	13.8	81.2	10978	15	AE012135	AE012135 Xanthomon
C 837	13.8	81.2	8454	2	CQ598358	CQ598358 Sequence	C 910	13.8	81.2	11006	2	CQ789654	CQ789654 Sequence
838	13.8	81.2	8540	2	AR198722	AR198722 Sequence	C 911	13.8	81.2	11059	2	CQ789683	CQ789683 Sequence
C 839	13.8	81.2	8588	10	AF257515	AF257515 Bovine le	C 912	13.8	81.2	11242	15	AE011933	AE011933 Xanthomon
C 840	13.8	81.2	8742	2	AX553924	AX553924 Sequence	C 913	13.8	81.2	11407	15	AE004775	AE004775 Pseudomon
841	13.8	81.2	8815	4	SMSMTA	Y10616 S macrospor	C 914	13.8	81.2	11646	2	AX577790	AX577790 Sequence
842	13.8	81.2	9027	2	AX577791	AX577791 Sequence	C 915	13.8	81.2	12041	2	AX577789	AX577789 Sequence
843	13.8	81.2	9320	2	AX663075	AX663075 Sequence	C 916	13.8	81.2	12055	15	AE009467	AE009467 Brucella
844	13.8	81.2	9382	10	SHIULGNS	L00676 Pseudorabie	C 917	13.8	81.2	12292	15	AE009033	AE009033 Agrobacte
845	13.8	81.2	9482	2	CQ829527	CQ829527 Sequence	C 918	13.8	81.2	12378	15	AE004797	AE004797 Pseudomon
846	13.8	81.2	9482	2	CS000580	CS000580 Sequence	C 919	13.8	81.2	12575	15	AE007999	AE007999 Agrobacte
847	13.8	81.2	9737	2	AR215118	AR215118 Sequence	C 920	13.8	81.2	12753	15	EBWCETS	EB7280 Pantoea agg
848	13.8	81.2	9737	2	AR215119	AR215119 Sequence	C 921	13.8	81.2	12840	8	DQ322638	DQ322638 VEEV repl
849	13.8	81.2	9737	2	AR215124	AR215124 Sequence	C 922	13.8	81.2	12840	8	DQ322640	DQ322640 VEEV repl
850	13.8	81.2	9737	2	AR302360	AR302360 Sequence	C 923	13.8	81.2	12840	8	DQ322642	DQ322642 VEEV repl
851	13.8	81.2	9737	2	AR302361	AR302361 Sequence	C 924	13.8	81.2	13068	2	CQ790429	CQ790429 Sequence
852	13.8	81.2	9737	2	AR302366	AR302366 Sequence	C 925	13.8	81.2	13547	2	AX577782	AX577782 Sequence
853	13.8	81.2	9737	2	AR373232	AR373232 Sequence	C 926	13.8	81.2	13766	2	AX553919	AX553919 Sequence
854	13.8	81.2	9737	2	AR373233	AR373233 Sequence	C 927	13.8	81.2	14136	11	TRU544612	AU544612 Takifugu
855	13.8	81.2	9737	2	AR373238	AR373238 Sequence	C 928	13.8	81.2	14338	15	AE010303	AE010303 Mechano
856	13.8	81.2	9737	2	AR401618	AR401618 Sequence	C 929	13.8	81.2	14818	15	AE010140	AE010140 Pyrococcu
857	13.8	81.2	9737	2	AR401619	AR401619 Sequence	C 930	13.8	81.2	14863	8	AY339820	AY339820 Nx-Lox ta
858	13.8	81.2	9737	2	AR401624	AR401624 Sequence	C 931	13.8	81.2	14987	2	CS058097	CS058097 Sequence
859	13.8	81.2	9737	2	AR442951	AR442951 Sequence	C 932	13.8	81.2	15081	2	CQ790448	CQ790448 Sequence
860	13.8	81.2	9737	2	AR442952	AR442952 Sequence	C 933	13.8	81.2	15083	2	CS101424	CS101424 Sequence
861	13.8	81.2	9737	2	AR442957	AR442957 Sequence	C 934	13.8	81.2	15110	2	CS101426	CS101426 Sequence
862	13.8	81.2	9737	2	AR568593	AR568593 Sequence	C 935	13.8	81.2	15407	11	AF536192	AF536192 Gallus ga
863	13.8	81.2	9737	2	AR568594	AR568594 Sequence	C 936	13.8	81.2	17290	2	AX708622	AX708622 Sequence
864	13.8	81.2	9737	2	AR568599	AR568599 Sequence	C 937	13.8	81.2	17344	2	CQ584942	CQ584942 Sequence
865	13.8	81.2	9737	2	AR568614	AR568614 Sequence	C 938	13.8	81.2	17358	2	CQ584942	CQ584942 Sequence
866	13.8	81.2	9737	2	AR568615	AR568615 Sequence	C 939	13.8	81.2	17402	2	CS165133	CS165133 Sequence
867	13.8	81.2	9737	2	AR568620	AR568620 Sequence	C 940	13.8	81.2	17753	2	CQ790449	CQ790449 Sequence
C 868	13.8	81.2	9846	5	ALJ359506	ALJ359506 Human DNA	C 941	13.8	81.2	18116	2	CS165132	CS165132 Sequence
869	13.8	81.2	9871	2	AR215120	AR215120 Sequence	C 942	13.8	81.2	18365	11	AY138246	AY138246 Gallus su
870	13.8	81.2	9871	2	AR302362	AR302362 Sequence	C 943	13.8	81.2	18495	15	BSU93875	BSU93875 Bacillus su
871	13.8	81.2	9871	2	AR373234	AR373234 Sequence	C 944	13.8	81.2	20216	2	CQ363808	CQ363808 Sequence
872	13.8	81.2	9871	2	AR401620	AR401620 Sequence	C 945	13.8	81.2	22183	12	AC014957	AC014957 Drosophila
873	13.8	81.2	9871	2	AR442953	AR442953 Sequence	C 946	13.8	81.2	22478	15	AB113568	AB113568 Actinopia
874	13.8	81.2	9871	2	AR568595	AR568595 Sequence	C 947	13.8	81.2	24102	12	AC018225	AC018225 Drosophila
875	13.8	81.2	9871	2	AR568596	AR568596 Sequence	C 948	13.8	81.2	25760	2	AX708628	AX708628 Sequence
876	13.8	81.2	10060	2	AR215121	AR215121 Sequence	C 949	13.8	81.2	26550	5	AF421855	AF421855 Homo sapi
877	13.8	81.2	10060	2	AR302363	AR302363 Sequence	C 950	13.8	81.2	27251	5	BX284661	BX284661 Human DNA
878	13.8	81.2	10060	2	AR373235	AR373235 Sequence	C 951	13.8	81.2	27435	2	CQ873891	CQ873891 Sequence
879	13.8	81.2	10060	2	AR401621	AR401621 Sequence	C 952	13.8	81.2	27435	5	AF535142	AF535142 Homo sapi
880	13.8	81.2	10060	2	AR442954	AR442954 Sequence	C 953	13.8	81.2	27652	5	AF495910	AF495910 Homo sapi
881	13.8	81.2	10060	2	AR568596	AR568596 Sequence	C 954	13.8	81.2	28058	2	AR619857	AR619857 Sequence
882	13.8	81.2	10060	2	AR568617	AR568617 Sequence	C 955	13.8	81.2	28762	2	AR619837	AR619837 Sequence
883	13.8	81.2	10188	15	AE012274	AE012274 Xanthomon	C 956	13.8	81.2	29183	12	AC014153	AC014153 Drosophila
884	13.8	81.2	10222	6	D89801	D89801 Mus musculu	C 957	13.8	81.2	29793	5	DQ173606	DQ173606 Homo sapi
885	13.8	81.2	10285	2	AX551315	AX551315 Sequence	C 958	13.8	81.2	29899	2	AR619870	AR619870 Sequence
886	13.8	81.2	10285	2	AX552015	AX552015 Sequence	C 959	13.8	81.2	30464	5	DQ173597	DQ173597 Homo sapi
C 887	13.8	81.2	10286	15	AE012336	AE012336 Xanthomon	C 960	13.8	81.2	32846	2	AX647129	AX647129 Sequence
888	13.8	81.2	10325	2	AR210352	AR210352 Sequence	C 961	13.8	81.2	32972	5	DQ173593	DQ173593 Homo sapi
889	13.8	81.2	10330	2	CQ789661	CQ789661 Sequence	C 962	13.8	81.2	32972	5	DQ173598	DQ173598 Homo sapi
890	13.8	81.2	10477	2	CQ789659	CQ789659 Sequence	C 963	13.8	81.2	32972	5	DQ173611	DQ173611 Homo sapi
891	13.8	81.2	10516	2	CQ789657	CQ789657 Sequence	C 964	13.8	81.2	32975	5	DQ173592	DQ173592 Homo sapi
892	13.8	81.2	10551	2	AX577781	AX577781 Sequence	C 965	13.8	81.2	32975	5	DQ173594	DQ173594 Homo sapi
893	13.8	81.2	10561	2	CQ789655	CQ789655 Sequence	C 966	13.8	81.2	32975	5	DQ173595	DQ173595 Homo sapi
894	13.8	81.2	10615	2	CQ789682	CQ789682 Sequence	C 967	13.8	81.2	32975	5	DQ173596	DQ173596 Homo sapi

968 13.8 81.2 32975 5 DQ173599 Homo sapi  
969 13.8 81.2 32975 5 DQ173600 Homo sapi  
970 13.8 81.2 32975 5 DQ173602 Homo sapi  
971 13.8 81.2 32975 5 DQ173603 Homo sapi  
972 13.8 81.2 32975 5 DQ173604 Homo sapi  
973 13.8 81.2 32975 5 DQ173605 Homo sapi  
974 13.8 81.2 32975 5 DQ173607 Homo sapi  
975 13.8 81.2 32975 5 DQ173608 Homo sapi  
976 13.8 81.2 32975 5 DQ173609 Homo sapi  
977 13.8 81.2 32975 5 DQ173610 Homo sapi  
978 13.8 81.2 32975 5 DQ173612 Homo sapi  
979 13.8 81.2 32975 5 DQ173613 Homo sapi  
980 13.8 81.2 32975 5 DQ173614 Homo sapi  
981 13.8 81.2 32975 5 DQ173615 Homo sapi  
982 13.8 81.2 32975 5 DQ173616 Homo sapi  
983 13.8 81.2 32975 5 DQ173618 Homo sapi  
984 13.8 81.2 32975 5 DQ173619 Homo sapi  
985 13.8 81.2 32975 5 DQ173620 Homo sapi  
986 13.8 81.2 32975 5 DQ173621 Homo sapi  
987 13.8 81.2 32975 5 DQ173623 Homo sapi  
988 13.8 81.2 32975 5 DQ173624 Homo sapi  
989 13.8 81.2 32975 5 DQ173625 Homo sapi  
990 13.8 81.2 32975 5 DQ173626 Homo sapi  
991 13.8 81.2 32975 5 DQ173627 Homo sapi  
992 13.8 81.2 32975 5 DQ173628 Homo sapi  
993 13.8 81.2 32975 5 DQ173629 Homo sapi  
994 13.8 81.2 32975 5 DQ173630 Homo sapi  
995 13.8 81.2 32975 5 DQ173631 Homo sapi  
996 13.8 81.2 32975 5 DQ173632 Homo sapi  
997 13.8 81.2 32975 5 DQ173633 Homo sapi  
998 13.8 81.2 32975 5 DQ173634 Homo sapi  
999 13.8 81.2 32975 5 DQ173635 Homo sapi  
1000 13.8 81.2 32975 5 DQ173636 Homo sapi

## ALIGNMENTS

RESULT 1  
BD233463 603 bp DNA linear PAT 17-JUL-2003  
LOCUS Human protein having hydrophobic domain and DNA encoding the same.  
DEFINITION  
ACCESSION BD233463  
VERSION BD233463.1 GI:33043233  
KEYWORDS JP 2002519016-A/9.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
REFERENCE 1 (bases 1 to 603)  
AUTHORS Kato,S. and Kimura,T.  
TITLE Human protein having hydrophobic domain and DNA encoding the same  
JOURNAL Patent: JP 2002519016-A 9 02-JUL-2002;  
SAGAMI CHEMICAL RESEARCH CENTER, PROTEGENE INC  
COMMENT OS Homo sapiens (human)  
PN JP 2002519016-A/9  
PD 02-JUL-2002  
PF 18-JUN-1999 JP 2000557267  
PI SEISHI KATO,TOMOKO KIMURA  
PC C12N15/09,C07K14/47,C12N1/15,C12N1/19,C12N5/10,C12N5/00,C12N5/ PC  
00 Human protein having hydrophobic domain and DNA encoding the  
CC Human protein having hydrophobic domain and DNA encoding the  
FH same  
FT Key Location/Qualifiers  
FT source 1..603  
/organism='Homo sapiens (human)'.  
LOCATION/Qualifiers  
1..603  
/organism='Homo sapiens'  
/mol\_type='genomic DNA'  
/db\_xref='taxon:9606'

ORIGIN  
Query Match 100.0%; Score 17; DB 2; Length 603;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GCCCAGCGTTGGCCGAG 17  
|||||  
Db 550 GCCCAGCGTTGGCCGAG 566  
RESULT 2  
BD209699 1212 bp DNA linear PAT 17-JUL-2003  
LOCUS Compositions isolated from skin cells and methods for their use.  
DEFINITION  
ACCESSION BD209699  
VERSION BD209699.1 GI:33019469  
KEYWORDS JP 2002512798-A/171.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1212)  
AUTHORS Strachan,L., Sleeman,M., Watson,J.D., Onrust,R., Kumble,A. and  
Murison,J.G.  
TITLE Compositions isolated from skin cells and methods for their use  
JOURNAL Patent: JP 2002512798-A 171 08-MAY-2002;  
GENESIS RESEARCH AND DEVELOPMENT CORP LTD  
COMMENT OS Homo sapiens (human)  
PN JP 2002512798-A/171  
PD 08-MAY-2002  
PF 29-APR-1999 JP 2000546009  
PR 29-APR-1998 US 09/069726,09-NOV-1998 US 09/188930 PI  
LORNA STRACHAN,MATTHEW SLEEMAN,JAMES DOUGLAS WATSON,RENE PI  
ONRUST.  
PI ANAND KUMBLE,JAMES GREG MURISON  
PC C12N15/09,A61K38/00,A61P9/00,A61P17/00,A61P29/00,A61P31/18, PC  
A61P35/00,  
PC C07K14/47,C12N1/15,C12N1/19,C12N1/21,C12N5/10,C12P21/02,C12N15/ PC  
00,  
FH A61K37/02,C12N5/00  
CC Compositions isolated from skin cells and methods for their  
CC use.  
FH Key Location/Qualifiers  
FT source 1..1212  
/organism='Homo sapiens (human)'.  
FEATURES  
source  
1..1212  
/organism='Homo sapiens'  
/mol\_type='genomic DNA'  
/db\_xref='taxon:9606'

ORIGIN  
Query Match 100.0%; Score 17; DB 2; Length 1212;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GCCCAGCGTTGGCCGAG 17  
|||||  
Db 990 GCCCAGCGTTGGCCGAG 1006  
RESULT 3  
AR341505 1212 bp DNA linear PAT 17-AUG-2003  
LOCUS Sequence 249 from patent US 6573095.  
DEFINITION  
ACCESSION AR341505  
VERSION AR341505.1 GI:33733640  
KEYWORDS .  
SOURCE Unknown.  
ORGANISM Unknown.

Unclassified.  
REFERENCE 1 (bases 1 to 1212)  
AUTHORS Strachan,L.  
TITLE Polynucleotides isolated from skin cells  
JOURNAL Patent: US 6573095-A 249 03-JUN-2003;  
Genesis Research & Development Corporation Limited, Parnell,  
NZX;  
FEATURES Location/Qualifiers  
source 1..1212  
/organism="unknown"  
/mol\_type="genomic DNA"  
ORIGIN  
Query Match 100.0%; Score 17; DB 2; Length 1212;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GCCCAGCTTGGCCGAG 17  
|||||  
Db 990 GCCCAGCTTGGCCGAG 1006  
RESULT 4  
AX078375  
LOCUS AX078375 1619 bp DNA linear PAT 22-FEB-2001  
DEFINITION Sequence 43 from Patent WO0107612.  
ACCESSION AX078375  
VERSION AX078375.1 GI:13158044  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
REFERENCE 1  
AUTHORS Au-Young,J., Bandman,O., Tang,Y.T., Yue,H., Azimzai,Y., Burford,N.,  
Baughn,M.R., Lu,D.A., Hillman,J.L., Patterson,C. and Lal,P.  
TITLE Receptors and associated proteins  
JOURNAL Patent: WO 0107612-A 43 01-FEB-2001;  
Incyte Genomics, Inc. (US)  
FEATURES Location/Qualifiers  
source 1..1619  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
/note="Incyte ID No: 2681738CB1"  
ORIGIN  
Query Match 100.0%; Score 17; DB 2; Length 1619;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GCCCAGCTTGGCCGAG 17  
|||||  
Db 1137 GCCCAGCTTGGCCGAG 1153  
RESULT 5  
BD233473  
LOCUS BD233473 1718 bp DNA linear PAT 17-JUL-2003  
DEFINITION Human protein having hydrophobic domain and DNA encoding the same.  
ACCESSION BD233473  
VERSION BD233473.1 GI:33043243  
KEYWORDS JP 2002519016-A/19.  
SOURCE Homo sapiens (human)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1718)  
AUTHORS Kato,S. and Kimura,T.  
TITLE Human protein having hydrophobic domain and DNA encoding the same  
JOURNAL Patent: JP 2002519016-A 19 02-JUL-2002;

SAGAMI CHEMICAL RESEARCH CENTER, PROTEGENE INC  
OS Homo sapiens (human)  
PN JP 2002519016-A/19  
PD 02-JUL-2002  
PF 18-JUN-1999 JP 2000557267  
PI SEISHI KATO,TOMOKO KIMURA  
PC  
C12N15/09,C07K14/47,C12N1/15,C12N1/19,C12N5/10,C12N15/00,C12N5/ PC  
00  
CC Human protein having hydrophobic domain and DNA encoding the  
CC same  
FH Key 1..1718  
FT source /organism="Homo sapiens (human)".  
FEATURES Location/Qualifiers  
source 1..1718  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Query Match 100.0%; Score 17; DB 2; Length 1718;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GCCCAGCTTGGCCGAG 17  
|||||  
Db 561 GCCCAGCTTGGCCGAG 577  
RESULT 6  
AK172760  
LOCUS AK172760 1826 bp mRNA linear PRI 07-MAY-2004  
DEFINITION Homo sapiens cDNA FLJ23921 fis, clone COL02043, highly similar to  
Homo sapiens retinoic acid induced 3 (RAI3).  
ACCESSION AK172760  
VERSION AK172760.1 GI:47077732  
KEYWORDS oligo capping: fis (full insert sequence).  
SOURCE Homo sapiens (human)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
REFERENCE 1  
AUTHORS Kawabata,A., Hikiiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,  
Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,  
Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.  
TITLE NEDO human cDNA sequencing project  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 1826)  
AUTHORS Sugano,S. and Suzuki,Y.  
TITLE Direct Submission  
JOURNAL Submitted (22-APR-2004) Sumio Sugano, Institute of Medical Science,  
University of Tokyo, Laboratory of Genome Structure, Human Genome  
Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan  
(E-mail:flcdna@mail.ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,  
Fax:81-3-5449-5416)  
COMMENT NEDO human cDNA sequencing project supported by Ministry of  
Economy, Trade and Industry of Japan; cDNA full insert sequencing;  
Research Association for Biotechnology; cDNA library construction;  
5'- & 3'-end one pass sequencing; Department of Virology and Human  
Genome Center, Institute of Medical Science, University of Tokyo  
(partly supported by Science and Technology Agency).  
FEATURES Location/Qualifiers  
source 1..1826  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="COL02043"  
/tissue\_type="colon"  
/clone\_lib="COL"  
/note="cloning vector pME18SFL3"  
ORIGIN

Query Match 100.0%; Score 17; DB 5; Length 1826;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACGCTTGCCGAG 17  
 |||||  
 DB 638 GCCACGCTTGCCGAG 654

RESULT 7  
 CQ723177  
 LOCUS CQ723177 2290 bp DNA linear PAT 03-FEB-2004  
 DEFINITION Sequence 9111 from Patent WO02068579.  
 ACCESSION CQ723177  
 VERSION CQ723177.1 GI:42284034  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homnidae; Homo.

REFERENCE 1  
 Venter, C.J., Adams M.C., Li, P.W. and Myers, E.W.  
 Kits, such as nucleic acid arrays, comprising a majority of  
 humanexons or transcripts, for detecting expression and other uses  
 thereof  
 Patent: WO 02068579-A 9111 06-SEP-2002;  
 PE Corporation (NY) (US)  
 Location/Qualifiers  
 1. .2290  
 /organism="Homo sapiens"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:9606"

FEATURES  
 source  
 ORIGIN  
 Query Match 100.0%; Score 17; DB 2; Length 2290;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACGCTTGCCGAG 17  
 |||||  
 DB 1117 GCCACGCTTGCCGAG 1133

RESULT 8  
 BC003665  
 LOCUS BC003665 2296 bp mRNA linear PRI 24-NOV-2004  
 DEFINITION Homo sapiens G protein-coupled receptor, family C, group 5, member  
 A, mRNA (cDNA clone MGC:923 IMAGE:2988011), complete cds.  
 ACCESSION BC003665  
 VERSION BC003665.2 GI:33872669  
 KEYWORDS MGC.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homnidae; Homo.

REFERENCE 1 (bases 1 to 2296)  
 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
 Klausner, R.D., Collins, P.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,  
 Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,  
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Heide, F.,  
 Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
 Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
 Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,  
 Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,  
 Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J.,  
 McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
 Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,  
 Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
 Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,  
 Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,

Buffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,  
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
 Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smallos, D.E.,  
 Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.  
 Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
 12477932  
 2 (bases 1 to 2296)  
 Director MGC Project.  
 Direct Submission  
 Submitted (26-FEB-2001) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 On Aug 19, 2003 this sequence version replaced gi:13177795.  
 Contact: MGC help desk  
 Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Institute for Systems Biology  
<http://www.systemsbio.org>  
 contact: [amadan@systemsbiology.org](mailto:amadan@systemsbiology.org)  
 Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha  
 Madan, Stephanie Rodriguez, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAL Plate: 3 Row: 1 Column: 11  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: matched mRNA gi: 12056470.  
 Location/Qualifiers  
 1. .2296  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="MGC:923 IMAGE:2988011"  
 /tissue\_type="Colon, adenocarcinoma"  
 /clone\_lib="NIH MGC\_15"  
 /lab\_host="DH10B-R"  
 /note="Vector: pOTB7"  
 1. .2296  
 /gene="GPCR5A"  
 /note="synonyms: RAIG1, GPRCSA, RAI3"  
 /db\_xref="GeneID:9052"  
 /db\_xref="MIM:604138"  
 95. .1168  
 /gene="GPCR5A"  
 /codon\_start=1  
 /product="G protein-coupled receptor, family C, group 5,  
 member A"  
 /protein\_id="AAH03665.1"  
 /db\_xref="GI:1317796"  
 /db\_xref="GeneID:9052"  
 /db\_xref="MIM:604138"  
 /translation="MATVPGCCNGLSKYIRLCDKRAANGIVLETVATAGVVTSA  
 FMTLPILVCKVQDSNRKMLPTQFLGLVIGLFGTLTFAIIGLDGSGTPRFFLFG  
 ILFSICFSLAHAVSLTKIVGRKPLSLVILGLAVGFSLVQDVIAIEYIVLTNRT  
 NVNVSLSAPRRNEDFVLLTYVFLMALFTLMSFTFCGSGFTGWRKHGARIYLTML  
 LSAIWAWITLLMLPDFDRWDOTILSSALAANGVFLLAYVSPFWLLTKQRNPD  
 YPVEDACKPOLVKYSGVENRAYSQBEITQFETGDTLYAVSYTHFQLQNPQPKQ  
 FSPRAHAWPSFYKDYEVKKEGS"

ORIGIN  
 Query Match 100.0%; Score 17; DB 5; Length 2296;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACGCTTGCCGAG 17  
 |||||  
 DB 1112 GCCACGCTTGCCGAG 1128

```
RESULT 9
AF506289
LOCUS Homo sapiens orphan G protein-coupling receptor PEIG-1 mRNA, PRI 15-JUL-2002
DEFINITION complete cds.
ACCESSION AF506289.1 GI:21779962
VERSION AF506289
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
REFERENCE
AUTHORS Terrett, J.A.
TITLE Diagnosis of carcinoma using raig1 polypeptides
JOURNAL Patent: WO 03087832-A 2 23-OCT-2003;
Oxford GlycoSciences (UK) Limited (GB)
FEATURES
source
1..2302
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 100.0%; Score 17; DB 2; Length 2302;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GCCACGCTTGGCCGAG 17
Db 1117 GCCACGCTTGGCCGAG 1133
RESULT 11
AF095448
LOCUS Homo sapiens putative G protein-coupled receptor (RAIG1) mRNA, PRI 29-DEC-1998
DEFINITION complete cds.
ACCESSION AF095448
VERSION AF095448.1 GI:4063889
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
REFERENCE
AUTHORS Cheng, Y. and Lotan, R.
TITLE Molecular cloning and characterization of a novel retinoic acid-inducible gene that encodes a putative G protein-coupled receptor
JOURNAL J. Biol. Chem. 273 (52), 35008-35015 (1998)
PUBMED 9857033
REFERENCE
AUTHORS Cheng, Y. and Lotan, R.
TITLE Direct Submision
JOURNAL Submitted (27-SEP-1998) Tumor Biology, The University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX 77030, USA
FEATURES
source
1..2302
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="12"
/map="12p12.3-p13"
1..2302
/gene="RAIG1"
/notes="induced by retinoic acid"
100..1173
/gene="RAIG1"
/codon_start=1
/product="putative G protein-coupled receptor"
/protein_id="AAC98506.1"
/db_xref="GI:4063890"
/translacion="MATTVPDGCNGLSKYYRLCDKAEANGIVLETATAGVVTSSA
FMTLPLVCKVQDSNRRKMLPTQFLFLGLVIGLFTFAFIIGLDGSGTPRFLFG
ILFSICFSLAHAVSLTKLVGRKRPUSLLVLGLAVGFSLVQDVIAIEYIVLTNNRT
NVNVSFELSAPRRNEDFVLLTYVFLMALTFMSSFTCGSGTGWKRHGAH1YLTML
LSIAIWMWITLLMPLDPDRWDITLSSALAANGWVFLAYVSPFWLLTKQRNPM
YPVDFACPKVKKGVENRAYSQBEITQTGFEETGDTLVAPYSTHFLQNQPQKE
FSIPRAHWPSPYKDYEVKKEGS"
ORIGIN
Query Match 100.0%; Score 17; DB 5; Length 2297;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GCCACGCTTGGCCGAG 17
Db 1125 GCCACGCTTGGCCGAG 1141
RESULT 10
AX930411
LOCUS Sequence 2 from Patent WO03087832.
DEFINITION AX930411
ACCESSION AX930411
VERSION AX930411.1 GI:40312209
```

LSTAIWVWITLLMLPDRRRDDTILSSALAANGWVFLLAYVSPFWLLTKQRNPMD  
YPVEDAFCKPQLVKYSYGVENRAYSQBEITQGPETGDTLYAPYSTHFLQNLQNPQKE  
FSIPRAHWPSPYKDYEVKKEGS"

## ORIGIN

Query Match 100.0%; Score 17; DB 5; Length 2302;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17

Db 1117 GCCCAGCGTTGGCCGAG 1133

## RESULT 12

BD156680

LOCUS BD156680 2446 bp DNA linear PAT 17-JAN-2003  
DEFINITION Primer for synthesizing full-length cDNA and use thereof.

ACCESSION BD156680

VERSION BD156680.1 GI:27862438

KEYWORDS JP 2002191363-A/11523.

SOURCE Homo sapiens (human)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.

1 (bases 1 to 2446)

Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,

Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.

Primer for synthesizing full-length cDNA and use thereof

Patent: JP 2002191363-A 11523 09-JUL-2002;

HELIX RESEARCH INSTITUTE

OS Homo sapiens (human)

PN JP 2002191363-A/11523

PD 09-JUL-2002

PF 28-JUL-2000 JP 2000280990

PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU

PI SAITO,

PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,

PI KEIICHI NAGAI, TETSUJI OTSUKI

## PC

C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC

10, C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC

Primer for synthesizing full-length cDNA and use thereof FH Key

Location/Qualifiers

FT CDS (254) . (1324) .

## FEATURES

source

1. .2446

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

## ORIGIN

Query Match 100.0%; Score 17; DB 2; Length 2446;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17

Db 1271 GCCCAGCGTTGGCCGAG 1287

## RESULT 13

AX877483

LOCUS AX877483 2446 bp DNA linear PAT 17-DEC-2003

DEFINITION Sequence 12388 from Patent EP1074617.

ACCESSION AX877483

VERSION AX877483.1 GI:40032219

KEYWORDS Homo sapiens (human)

## SOURCE

\*ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.

## REFERENCE

## AUTHORS

Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,  
Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.

Primers for synthesizing full-length cDNA and their use

Patent: EP 1074617-A 12388 07-FEB-2001;

Research Association for Biotechnology (JP)

## FEATURES

source

1. .2446

/organism="Homo sapiens"

/mol\_type="unassigned DNA"

/db\_xref="taxon:9606"

254. .1327

/note="unnamed protein product"

/codon\_start=1

/protein\_id="CAE89746.1"

/db\_xref="GI:40032220"

/translation="MATTVPDGCENGLKSKYVRLCDKAEAGIVLETVATAGVVTVA

FMULTPLVCKVQDSNRKMLPTQFLLGLVGLGFTAFIIGLDSGTGTRPFLFG

ILFSICFCLLAHAVSLTKLVGRKPLSLVILGLAVGFSLVQDVIAIEIVLTNNRT

NNVFSLSAPRRNEDEVLLTYVFLMALTFMSSFTFCGSGTWKRHGHAHYLTML

LSTAIWVWITLLMLPDRRRDDTILSSALAANGWVFLLAYVSPFWLLTKQRNPMD

YPVEDAFCKPQLVKYSYGVENRAYSQBEITQGPETGDTLYAPYSTHFLQNLQNPQKE

FSIPRAHWPSPYKDYEVKKEGS"

## ORIGIN

Query Match 100.0%; Score 17; DB 2; Length 2446;

Best Local Similarity 100.0%; Pred. No. 1.5e+02;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17

Db 1271 GCCCAGCGTTGGCCGAG 1287

## RESULT 14

AK001761

LOCUS AK001761

DEFINITION Homo sapiens cDNA FLJ10899 fis, clone NT2RP5003506.

ACCESSION AK001761

VERSION AK001761.1 GI:7023229

KEYWORDS oligo capping; fis (full insert sequence).

SOURCE Homo sapiens (human)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.

## REFERENCE

## AUTHORS

Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R.,

Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H.,

Sekine, M., Obayashi, M., Nishi, T., Shibahara, T., Tanaka, T.,

Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y.,

Nagahari, K., Murakami, K., Yasuda, T., Iwayanagi, T., Wagatsuma, M.,

Shiratori, A., Sudo, H., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H.,

Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T.,

Kikkawa, E., Omura, Y., Abe, K., Kamiyama, K., Katsuta, N., Sato, K.,

Tanikawa, M., Yamazaki, M., Ninomiya, K., Ishibashi, T., Yamashita, H.,

Murakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M.,

Hiraoka, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S.,

Yosida, M., Hotuta, T., Kusano, J., Kanehori, K., Takahashi-Fujii, A.,

Hara, H., Tanase, T.O., Nomura, Y., Togiya, S., Komai, F., Hara, R.,

Takeuchi, K., Arita, M., Imose, N., Musashino, K., Yuuki, H., Oshima, A.,

Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T.,

Shiohara, N., Sano, S., Moriya, S., Momiyama, H., Satoh, N., Takami, S.,

Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H.,

Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T.,

Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K.,

Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M.,

Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yamada, K., Fujii, Y.,

Ozaki, K., Hirao, M., Ohmori, Y., Kawabata, A., Hikiji, T., Kobatake, N.,

Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Kawakami, T.,

Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsumura, K.,

Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T., Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J., Satoh, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K., Nagase, T., Nomura, N., Kikuchi, H., Masuho, Y., Yamashita, R., Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T., and Sugano, S. Complete sequencing and characterization of 21,243 full-length human cDNAs  
Nat. Genet. 36 (1), 40-45 (2004)  
14702039

Isogai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y., Nishikawa, T., Nagai, K., Sugano, S., Ishibashi, T., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M., Hiraoka, S., Ishii, S., Kawai, Y., Saito, K., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagahara, K., Masuho, Y., and Kanehori, K.  
NEDO human cDNA sequencing project  
Unpublished

3 (bases 1 to 2446)  
Isogai, T. and Otsuki, T.  
Direct Submission  
Submitted (16-FEB-2000) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana Kisarazu, Chiba 292-0812, Japan (E-mail: flj-cdn@nifty.com, tel: 81-438-52-3975, fax: 81-438-52-3986)  
NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert construction; Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing and clone selection; Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.  
Location/Qualifiers  
1. .2446  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="NT2RP5003506"  
/cell\_line="NT2"  
/cell\_type="teratocarcinoma"  
/clone\_lib="NT2RP5"  
/note="Cloning vector: pME18SFL3  
mRNA from NT2 neuronal precursor cells after 2-weeks retinoic acid (RA) induction."  
254. .1327  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="BAA91890.1"  
/db\_xref="GI:7023230"  
/translation="MATTPDGCRLGKSKYYRLCDKAEAWGIYLETVATAGVVTVA  
FMFLPILVCKVQDSNRKMLPTQFLFLGLVIGLTFAPFIIGLDGSGTRFFLFG  
ILFSICFSLAHAVSLTKLVGRKPLSLVLGLAVGFSLVQDVIAIEYIVLTNNRT  
NNVFSLSAPRRNEFVLLITVYVLEFALMTFLMSFTFCGSGFTGWRHGAHIVLTWL  
LSIAIYAWITLLMLPDRRWDITLSSALAANGWFFLLAYVSPFWLLTKQRNPD  
YPVEDACFKQLVKKSIGVENRAYSQEITGFEETGDTLYASTHTFQLQNQPQKE  
FSIPRAHAWSPYKDYEVKKEGS"

ORIGIN  
Query Match 100.0%; Score 17; DB 5; Length 2446;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGCCGAG 17  
|||||  
Db 1271 GCCCAGCGTTGCCGAG 1287

RESULT 15  
CQ981495  
LOCUS 2456 bp DNA linear PAT 25-JAN-2005  
DEFINITION Sequence 350 from Patent EP1498424.  
ACCESSION CQ981495  
VERSION CQ981495.1 GI:58190785  
KEYWORDS  
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1  
AUTHORS Rosenthal, A., Hermann, K., Heiden, E., Pilarsky, C., Bruemendorf, T., Staub, E., Roepcke, S., Mennerich, D., Kinnemann, H. and Li, X.  
TITLE Human nucleic acid sequences from lung tumours  
JOURNAL Patent: EP 1498424-A 350 19-JAN-2005;  
Hinzmann, Bernd (DE); Hermann, Klaus (DE); Heiden, Esmeralda (DE); Rosenthal, Andre (DE)  
FEATURES  
source  
1. .2456  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

ORIGIN  
Query Match 100.0%; Score 17; DB 2; Length 2456;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGCCGAG 17  
|||||  
Db 1271 GCCCAGCGTTGCCGAG 1287

RESULT 16  
DD210040  
LOCUS 2456 bp DNA linear PAT 19-JAN-2006  
DEFINITION Methods of Diagnosis of Cancer, Compositions and Methods of Screening for Modulators of Cancer.  
ACCESSION DD210040  
VERSION DD210040.1 GI:85654022  
KEYWORDS JP 2005518782-A/34.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 2456)  
Zlotnik, A., Mack, D.H., Agiz, N., Gish, K.C., Hebeji, P.A., Wilson, K.E. and Afar, D.  
TITLE Methods of Diagnosis of Cancer, Compositions and Methods of Screening for Modulators of Cancer  
JOURNAL Patent: JP 2005518782-A 34 30-JUN-2005;  
COMMENT PROTEIN DESIGN LABS INC  
OS Homo sapiens  
PN JP 2005518782-A/34  
PD 30-JUN-2005  
PF 17-SEP-2002 JP 2003529912  
PR 12-APR-2002 US 60/372246, 08-FEB-2002 US 60/355257, PR 08-FEB-2002 US 60/355145, 13-NOV-2001 US 60/350666, PR 20-SEP-2001 US 60/323887, 17-SEP-2001 US 60/323469 PI albert zlotnik, david h mack, natasha agiz, kurt c gish, peter a pi hebeji, PI keith e wilson, daniel afar  
CC  
FH  
Location/Qualifiers  
1. .2456  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

ORIGIN  
Query Match 100.0%; Score 17; DB 2; Length 2456;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGCCGAG 17  
|||||  
Db 1271 GCCCAGCGTTGCCGAG 1287

```

RESULT 17
AX549168
LOCUS
DEFINITION Sequence 453 from Patent WO02061087.
ACCESSION AX549168
VERSION AX549168.1 GI:25813894
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1
AUTHORS
TITLE Antigenic peptides, such as for G protein-coupled receptors
(GPCRs), antibodies thereto, and systems for identifying such
antigenic peptides
JOURNAL Patent: WO 02061087-A 453 08-AUG-2002;
Lifespan Biosciences, Inc. (US)
FEATURES
source
1. .2456
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 100.0%; Score 17; DB 2; Length 2456;
Best Local Similarity 100.0%; Pred. No. 1.5e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 17; Conservative 0;
Oy 1 GCCCAGCGTTGGCCGAG 17
|||||
Db 1271 GCCCAGCGTTGGCCGAG 1287

RESULT 18
AK122672
LOCUS
DEFINITION Homo sapiens cDNA FLJ16117 fis, clone ASTRO2003632, highly similar
to Homo sapiens putative G protein-coupled receptor (RAIG1) mRNA.
ACCESSION AK122672
VERSION AK122672.1 GI:34527861
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1
AUTHORS
Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R.,
Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H.,
Sekine, M., Obayashi, M., Nishi, T., Shibahara, T., Tanaka, T.,
Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y.,
Nagahari, K., Murakami, K., Yasuda, T., Iwayanagi, T., Wagatsuma, M.,
Shiratori, A., Sudo, H., Hosioki, T., Kaku, Y., Kodaira, H., Kondo, H.,
Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T.,
Kikkawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K.,
Tanikawa, M., Yanazaki, M., Nishimura, K., Ishibashi, T., Yamashita, H.,
Murakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M.,
Hiraoka, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S.,
Yosida, M., Hotta, T., Kusano, J., Kanehori, K., Kanehori, K., Tanabe, A.,
Hara, H., Tanase, T. O., Nomura, Y., Togiya, S., Komai, F., Hara, R.,
Takeuchi, K., Arita, M., Imose, N., Musashino, K., Yuuki, H., Oshima, A.,
Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T.,
Shiohata, N., Sano, S., Moriya, S., Momiyama, H., Satoh, N., Takami, S.,
Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H.,
Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T.,
Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K.,
Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M.,
Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yamada, K., Fujii, Y.,
Ozaki, K., Hirao, M., Ohmori, Y., Kawabata, A., Hikiji, T., Kobatake, N.,
Inagaki, H., Ikeda, Y., Okamoto, S., Okitani, R., Kawakami, T.,

```

```

Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsumura, K.,
Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togaishi, T.,
Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J.,
Sato, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K.,
Nagase, T., Nomura, N., Kikuchi, H., Masuho, Y., Yamashita, R.,
Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S.
Complete sequencing and characterization of 21,243 full-length
human cDNAs
Nat. Genet. 36 (1), 40-45 (2004)
JOURNAL
PUBMED 14702039
REFERENCE
2
AUTHORS
Tanigami, A., Fujiwara, T., Shibahara, T., Goto, Y., Hirao, M.,
Shimizu, F., Wakebe, H., Ono, T., Hishigaki, H., Watanabe, T., Ozaki, K.,
Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S.,
Yamamoto, J., Isono, Y., Kawai, H., Matsuo, K., Saito, K., Nishikawa, T.,
Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M.,
Kikuchi, H., Kanda, K., Wagatsuma, M., Murakawa, K., Kanehori, K.,
Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y.,
Sugano, S., Nagahari, K., Masuho, Y., Nagai, K. and Isogai, T.
NEDO human cDNA sequencing project
Unpublished
3 (bases 1 to 3057)
Isogai, T. and Yamamoto, J.
Direct Submission
Submitted (15-JUL-2003) Takao Isogai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail: flj-cdna@nifty.com, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: Reverse Proteomics Research Institute, HRI and
RAB.
FEATURES
source
Location/Qualifiers
1. .3057
/organism="Homo sapiens"
/mol_type="mrna"
/db_xref="taxon:9606"
/clone="ASTRO2003632"
/cell_type="normal astrocytes (NHA5732)"
/clone_lib="ASTRO2"
/note="cloning vector: pME18SFL3
primary culture, normal astrocytes"
ORIGIN
Query Match 100.0%; Score 17; DB 5; Length 3057;
Best Local Similarity 100.0%; Pred. No. 1.5e+02; Mismatches 0; Gaps 0;
Matches 17; Conservative 0;
Oy 1 GCCCAGCGTTGGCCGAG 17
|||||
Db 1908 GCCCAGCGTTGGCCGAG 1924

RESULT 19
AX188348
LOCUS
DEFINITION Sequence 4043 from Patent WO0142467.
ACCESSION AX188348
VERSION AX188348.1 GI:15139821
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1
AUTHORS
Schlegel, R., Deeds, J., Berger, A. and Zhao, X.
Genes, compositions, kits, and methods for identification,
assessment, prevention, and therapy of cervical cancer

```



JOURNAL Patent: WO 0142467-A 4043 14-JUN-2001;  
Millennium Predictive Medicine, Inc. (US)  
FEATURES Location/Qualifiers  
source 1..3371  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Query Match 100.0%; Score 17; DB 2; Length 3371;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GCCACGCTTGCCGAG 17  
|||||  
Db 1184 GCCACGCTTGCCGAG 1200  
RESULT 20  
CQ894732 6730 bp DNA linear PAT 05-NOV-2004  
LOCUS  
DEFINITION Sequence 42 from Patent EP1471075.  
ACCESSION CQ894732  
VERSION CQ894732.1 GI:55467481  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
REFERENCE 1  
AUTHORS Rosenthal,A.D., Pilarsky,C., Dahl,E., Specht,T., Bruemendorf,T.,  
Lichtner,R., Staub,E., Roepcke,S. and Li,X.I.  
TITLE Human nucleic acid sequences expressed in pancreatic carcinomas  
JOURNAL Patent: EP 1471075-A 42 27-OCT-2004;  
Hinzmann, Bernd (DE); Rosenthal, Andre (DE); Pilarsky, Christian  
(DE); Dahl, Edgar (DE); Specht, Thomas (DE); Lichtner, Rosemarie  
(DE)  
FEATURES Location/Qualifiers  
source 1..6730  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Query Match 100.0%; Score 17; DB 2; Length 6730;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GCCACGCTTGCCGAG 17  
|||||  
Db 1271 GCCACGCTTGCCGAG 1287  
RESULT 21  
AC007688/c 161577 bp DNA linear PRI 30-AUG-2002  
LOCUS  
DEFINITION Homo sapiens 12 BAC RPC111-392P7 (Roswell Park Cancer Institute  
Human BAC Library) complete sequence.  
ACCESSION AC007688  
VERSION AC007688.15 GI:5815499  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
REFERENCE 1 (bases 1 to 161577)  
AUTHORS Muzny,D., Aronson,A.D., Bouck,J., Bunac,C., Chen,Z., Ding,Y.,  
Dugan,S., Durbin,J., Forcum,J., Garcia,C., Gorrell,J.H.,  
Gorrell,L.L., Hernandez,J., Issar,A., Jackson,D., Kneitz,S.,  
Kondejewski,N., Lau,S., Leal,B., Lee,E., Lichtarge,O., Liu,W.,  
Logan,O., Lu,J., Marondel,I., Martinez,C., Merscher,S., Miller,A.,  
Montgomery,K., Oswal,G., Pampell,L.R., Parish,B.J., Perez,L.,  
Rashid,N.D., Rives,C., Scherer,S.E., Shen,H., Shim,C., Simon,M.,  
Vo,Q., Williamson,A., Worley,K.C., Xiang,A.M., Yang,R., Yu,W.,  
Zhou,X., Kucherlapati,R., Nelson,D. and Gibbs,R.A.  
Direct Submission  
Unpublished  
2 (bases 1 to 161577)  
Worley,K.C.  
Direct Submission  
Submitted (01-JUN-1999) Molecular and Human Genetics, Baylor  
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 161577)  
Worley,K.C.  
Direct Submission  
Submitted (01-SEP-1999) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
4 (bases 1 to 161577)  
Worley,K.C.  
Direct Submission  
Submitted (16-MAY-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
5 (bases 1 to 161577)  
Worley,K.C.  
Direct Submission  
Submitted (30-AUG-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Sep 1, 1999 this sequence version replaced gi:5757565.  
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email  
[gc-help@bcm.tmc.edu](mailto:gc-help@bcm.tmc.edu)  
CLONE LENGTH: This sequence does not necessarily represent the  
entire insert of this clone. Overlapping regions of clones are only  
sequenced and submitted once, so the sequence for the remainder of  
the insert may be found in the record for the adjacent clones.  
Overlapping clones are noted at the beginning and end of the  
Features listing.  
ANNOTATION OF FEATURES:  
STGs are identified using ePCR (Genome Res. 7:541-550) searches  
of a local database that includes entries from dbSTS, GDB, and  
local mapping efforts.  
Repeats are identified using RepeatMasker (A. Smit and P. Green,  
unpublished.) for Human and Mouse sequences.  
Genes and Region of sequence similarity are identified by BLAST  
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the  
EST and cDNA sequences. Genes demonstrate at least two exons  
flanked by consensus splice sites that maintained sequence  
continuity across the splice junctions. Sequences that are not  
identical matches are annotated as similar.  
SEQUENCING READ COVERAGE: Sequencing is completed to a minimum  
standard of double strand coverage with a minimum of 2 clones and 2  
reads with no ambiguities or 2 chemistries with a minimum of 2  
clones and 3 reads with no ambiguities. If the sequence quality for  
a region does not meet this standard, it will be indicated in the  
annotation as Low Coverage.  
QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality  
standards - estimated error rate less than 1 per 10,000 bases.  
Reports of lowest quality individual bases and measures of base  
quality are listed below. Description of the metrics can be found  
at URL:  
<http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.  
QUALSTAT-REPORT-----  
----- Summary Statistics -----  
Contig length: 161577  
Phrap values in estimate: 160751



Gaskin, C., Gench, S., Ghose, S., Gill, R., Gonzalez, D., Gonzalez-Garay, M., Guevara, W., Holder, M., Haaland, W., Haerberlen, K., Hall, B., Hamid, H., Hamilton, K., Harbes, B., Harris, R., Havlak, P., Hawes, A., Hawkins, E., Hayes, S., Hemphill, L., Hernandez, J., Hines, S., Hitchens, M., Hodgson, A., Hogue, M., Hollins, B., Howell, L. T., Hulyk, S., Hume, J., Imo, K., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Kalafas, K., Kelly, S., Keys, T., Khan, Z., King, L., Kovar, C., Kowis, A., Kowis, C., Lara, F., Leal, S., Lee, K., Lee, S., LeGall, F. I., Lemon, S., Lewis, L., Li, B., Li, Y., Li, Z., Linnell, M., Liu, W., Liu, Y.-S., Liu, Y., Liyanage, D., London, P., Lopez, J., Lorensuwa, L., Lozano, R., Luk, T., Madu, R., Maheshwari, M., Mahoney, C., Malloy, K., Mansouri, D., Martinez, E., McClelland, H., McPherson, J., Mercadao, C., Metzker, M., Milosavljevic, A., Minja, E., Morgan, M., Morris, S., Munidasa, M., Murray, D., Nazarith, L., Ngo, D., Nguyen, N., Norwig-Eastaugh, E., Nott, A., Nwaokeme, O., Oregon, M., Ochi-Okorie, C., Oden, E., Okwuonu, G., Okwuonu, K., Parker, D., Pasternak, S., Patel, B., Patel, V., Paul, H., Perez, A., Perez, L., Petrosino, J., Pham, T., Primus, E., Pu, L.-L., Puazo, M., Qin, X., Quinn, A., Quiroz, J., Rabata, D., Rachlin, E., Reich, R., Ren, Y., Reuter, M., Richards, S., Rives, C., Rodriguez, F., Rojas, A., Ruiz, S. J., Sana, M., Sanders, W., Santibanez, J., Santos, R., Savery, G., Scherer, S., Shen, H., Shen, Y., Sisson, I., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Svatek, A., Taylor, E., Taylor, T., Thomas, N., Thorn, R., Thornton, R., Trejos, Z., Usmani, K., Vargo, C., Verduzco, D., Villasana, D., Virk, D., Volkov, A., Waldron, L., Walker, B., Wang, O., Wang, S., Warren, J., Wei, X., Wheeler, D., Williams, G., Williams, R., Worley, K., Wright, R., Wu, J., Yakub, S., Yan, K., Yuan, Y., Yu, F., Zhang, J., Zhang, L., Zhang, Z., Zhou, J., Weinstein, G., and Gibbs, R. A.

## TITLE

Direct Submission

## REFERENCE

2 (bases 1 to 213601)

## AUTHORS

Worley, K. C.

## JOURNAL

Direct Submission

## REFERENCE

Submitted (02-JUL-2005) Human Genome Sequencing Center, Department

## AUTHORS

of Molecular and Human Genetics, Baylor College of Medicine, One

## JOURNAL

Baylor Plaza, Houston, TX 77030, USA

## REFERENCE

3 (bases 1 to 213601)

## AUTHORS

Bovine Genome Sequencing Consortium

## JOURNAL

Submitted (23-JAN-2006) Human Genome Sequencing Center, Department

## REFERENCE

of Molecular and Human Genetics, Baylor College of Medicine, One

## JOURNAL

Baylor Plaza, Houston, TX 77030, USA

## REFERENCE

On Jan 23, 2006 this sequence version replaced gi:68533315.

## AUTHORS

The sequence in this assembly is a combination of BAC based reads

## JOURNAL

and whole genome shotgun sequencing reads assembled using Atlas

## REFERENCE

(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described

## AUTHORS

in the feature table below represents a scaffold in the Atlas

## JOURNAL

assembly (a 'contig-scaffold'). Within each contig-scaffold

## REFERENCE

individual sequence contigs are ordered and oriented, and separated

## AUTHORS

by sized gaps filled with Ns to the estimated size. The sequence

## JOURNAL

may extend beyond the ends of the clone and there may be sequence

## REFERENCE

contigs within a contig-scaffold that consist entirely of whole

## AUTHORS

genome shotgun sequence reads. Both end sequences and whole genome

## JOURNAL

shotgun sequence only contigs will be indicated in the feature

## REFERENCE

table.

## AUTHORS

----- Genome Center

## JOURNAL

Center: Baylor College of Medicine

## REFERENCE

Center code: BCM

## AUTHORS

Web site: http://www.hgsc.bcm.tmc.edu/

## JOURNAL

Contact: hgsc-help@bcm.tmc.edu

## REFERENCE

----- Project Information

## AUTHORS

Center project name: FUGQ

## JOURNAL

Center clone name: CH240-159B22

## REFERENCE

----- Summary Statistics

## AUTHORS

Assembly program: Atlas 3.0;

## JOURNAL

Consensus quality: 204849 bases at least Q40

## REFERENCE

Consensus quality: 206683 bases at least Q30

## AUTHORS

Consensus quality: 208261 bases at least Q20

## JOURNAL

Estimated insert size: 212567; sum-of-contigs estimation

## REFERENCE

Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

## AUTHORS

-----

\* NOTE: Estimated insert size may differ from sequence length  
(see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 25 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 30391: contig of 30391 bp in length  
\* 30392 30441: gap of 50 bp  
\* 30442 32112: contig of 1671 bp in length  
\* 32113 32162: gap of 50 bp  
\* 32163 55102: contig of 22940 bp in length  
\* 55103 55188: gap of 86 bp  
\* 55189 57567: contig of 2379 bp in length  
\* 57568 57617: gap of 50 bp  
\* 57618 63739: contig of 6122 bp in length  
\* 63740 63789: gap of 50 bp  
\* 63790 81080: contig of 17291 bp in length  
\* 81081 81130: gap of 50 bp  
\* 81131 85683: contig of 4553 bp in length  
\* 85684 86125: gap of 442 bp  
\* 86126 108814: contig of 22689 bp in length  
\* 108815 108864: gap of 50 bp  
\* 108865 12979: contig of 18115 bp in length  
\* 12980 127029: gap of 50 bp  
\* 127030 135625: contig of 8596 bp in length  
\* 135626 135675: gap of 50 bp  
\* 135676 140495: contig of 4820 bp in length  
\* 140496 140545: gap of 50 bp  
\* 140546 145818: contig of 5273 bp in length  
\* 145819 145868: gap of 50 bp  
\* 145869 149186: contig of 3318 bp in length  
\* 149187 149236: gap of 50 bp  
\* 149237 156197: contig of 6961 bp in length  
\* 156198 156247: gap of 50 bp  
\* 156248 160593: contig of 4346 bp in length  
\* 160594 160643: gap of 50 bp  
\* 160644 168168: contig of 7525 bp in length  
\* 168169 168218: gap of 50 bp  
\* 168219 175899: contig of 7681 bp in length  
\* 175900 175949: gap of 50 bp  
\* 175950 177819: contig of 1870 bp in length  
\* 177820 177919: gap of unknown length  
\* 200888 200938: gap of 50 bp  
\* 200939 203883: contig of 2945 bp in length  
\* 203884 204161: gap of 278 bp  
\* 204162 208264: contig of 2103 bp in length  
\* 208265 208364: gap of unknown length  
\* 208365 207491: contig of 1127 bp in length  
\* 207492 207591: gap of unknown length  
\* 207592 208821: contig of 1230 bp in length  
\* 208822 208921: gap of unknown length  
\* 208922 210956: contig of 2035 bp in length  
\* 210957 211057: gap of unknown length  
\* 211057 213601: contig of 2545 bp in length.  
Location/Qualifiers  
1. 213601  
/organism="Bos taurus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9913"  
/clone="CH240-159B22"  
30392. .30441  
/estimated\_length=50  
32113. .32162  
/estimated\_length=50  
55103. .55188  
/estimated\_length=86  
57568. .57617  
/estimated\_length=50

FEATURES  
source

gap

gap

gap

gap

```

gap      63740..63789
/estimated length=50
gap      81081..81130
/estimated length=50
gap      85684..85725
/estimated length=442
gap      108815..108864
/estimated length=50
gap      126980..127029
/estimated length=50
gap      135626..135675
/estimated length=50
gap      140496..140545
/estimated length=50
gap      145819..145868
/estimated length=50
gap      149187..149236
/estimated length=50
gap      156198..156247
/estimated length=50
gap      160594..160643
/estimated length=50
gap      168169..168218
/estimated length=50
gap      175900..175949
/estimated length=50
gap      177820..177919
/estimated length=unknown
gap      200889..200938
/estimated length=50

Query Match      100.0%; Score 17; DB 12; Length 213601;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GCCACCGTTGCCGAG 17
        |||
Db      83186 GCCACCGTTGCCGAG 83170

RESULT 23
LOCUS   AC166087
DEFINITION Bos taurus clone CH240-103F13, *** SEQUENCING IN PROGRESS ***, 35
unordered pieces.
ACCESSION AC166087
VERSION   AC166087.2 GI:85664213
KEYWORDS  HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE    Bos taurus (cattle)
ORGANISM  Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
Pecora; Bovidae; Bovinae; Bos.
1 (bases 1 to 284106)
Muzny,D., Adams,C., Agbai II,O., Allen,C., Albrooks,S., Archer,P.,
Arredondo,H., Bandaranaike,D., Bangura,D., Beltran,B., Beltran,R.,
Beraducci,A., Biswal,K., Blyth,P., Bonham,H., Buhay,C., Burch,P.,
Cadoree,I., Canada,A., Cardenas,V., Carter,K., Cavazos,I.,
Chacko,J., Chahour,M., Chavez,D., Chen,A., Chen,G., Chen,R.,
Cheng,M.-T., Chu,J., Ciers,K., Cockrell,R., Coyle,M., Cree,A.,
Curry,S., Dai,W., Davila,M.L., Davis,C., Davy-Carroll,L., De
Anda,C., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H.,
Donlin,J., McCauley,S., Dugan-Rocha,S., Dunn,A., Durbin,K.,
Dziuda,D., Egan,A., Escotto,M., Espinosa,V., Eugene,C., Fa,M.,
Fernandez,S., Fernando,P., Flagg,N., Forbes,L., Foster,P.,
Fowler,G., Fu,Q., Fuh,E., Garcia,A., Garcia,R., Garner,T.,
Gaskin,C., Gench,S., Ghose,S., Gill,R., Gonzalez,D.,
Gonzalez-Garay,M., Guevara,W., Holder,M., Haaland,W., Haerberlen,K.,
Hall,B., Hamid,H., Hamilton,K., Harbes,B., Harris,R., Havlak,P.,
Hawes,A., Hawkins,E., Hayes,S., Hemphill,L., Hernandez,J.,
Hines,S., Hitchens,M., Hodgson,A., Hogues,M., Hollins,B.,
Howell,L.T., Hulyk,S., Hume,J., Ito,K., Jackson,A., Jackson,L.,
Jacob,L., Jiang,H., Johnson,B., Johnson,R., Kalafus,K., Kelly,S.,

```

```

Keys,T., Khan,Z., King,L., Kovar,C., Kowis,A., Kowis,C., Lara,F.,
Leal,S., Lee,K., Lee,S., LeGall,F.I., Lemon,S., Lewis,L., Li,B.,
Li,Y., Li,Z., Linell,M., Liu,W., Liu,Y.-S., Liu,Y., Liyangde,D.,
London,P., Lopez,J., Lorensuhera,L., Lozado,R., Luk,T., Madu,R.,
Maheshwari,M., Mahoney,C., Malloy,K., Mansouri,D., Martinez,E.,
McLelland,H., McPherson,J., Mercadao,C., Metzker,M.,
Milosavljevic,A., Minja,E., Morgan,M., Morris,S., Munidasa,M.,
Murray,D., Nazarith,L., Ngo,D., Nguyen,N., Norwig-Eastaugh,E.,
Nott,A., Nwaokemele,O., Obregon,M., Ochi-Okorie,C., Odeh,E.,
Okwuonu,G., Okwuonu,K., Parker,D., Pasternak,S., Patel,B.,
Patel,V., Paul,H., Perez,A., Perez,L., Petrosino,J., Pham,T.,
Primus,E., Pu,L.-L., Puazo,M., Qin,X., Quinn,A., Quiroz,J.,
Rabata,D., Rachlin,E., Reigh,R., Ren,Y., Reuter,M., Richards,S.,
Rives,C., Rodriguez,F., Rojas,A., Ruiz,S.J., Sana,M., Sanders,W.,
Santibanez,J., Santos,R., Savery,G., Scherer,S., Shen,H., Shen,Y.,
Sisson,I., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R.,
Svatek,A., Taylor,E., Taylor,T., Thomas,N., Thorn,R., Thornton,R.,
Trejos,Z., Usmani,K., Vargo,C., Verduzco,D., Villanana,D., Virk,D.,
Volkov,A., Waldron,L., Walker,B., Wang,Q.-Q., Wang,S., Warren,J.,
Wei,X., Wheeler,D., Williams,G., Williams,R., Worley,K., Wright,R.,
Wu,J., Yakub,S., Yan,K., Yuan,Y., Yu,F., Zhang,J., Zhang,L.,
Zhang,Z., Zhou,J., Weinstock,G. and Gibbs,R.A.

Direct Submission
Unpublished
2 (bases 1 to 284106)
Worley,K.C.

Direct Submission
Submitted (23-JUL-2005) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 284106)

Bovine Genome Sequencing Consortium
Direct Submission
Submitted (23-JAN-2006) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jan 23, 2006 this sequence version replaced gi:71067167.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu

----- Project Information
Center project name: FKHC
Center clone name: CH240-103F13
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 264399 bases at least Q40
Consensus quality: 267053 bases at least Q30
Consensus quality: 269711 bases at least Q20
Estimated insert size: 271703; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

-----
* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
* NOTE: this sequence may represent more than one 'clone'.
* NOTE: this is a 'working draft' sequence. It currently
consists of 35 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as

```

\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 3307: contig of 3307 bp in length  
3308 3357: gap of 50 bp  
3358 21882: contig of 18525 bp in length  
21883 21932: gap of 50 bp  
21933 27057: contig of 5125 bp in length  
27058 27107: gap of 50 bp  
27108 30386: contig of 3279 bp in length  
30387 30436: gap of 50 bp  
30437 34785: contig of 4349 bp in length  
34786 34912: gap of 127 bp  
34913 37140: contig of 2228 bp in length  
37141 37190: gap of 50 bp  
37191 50210: contig of 13020 bp in length  
50211 50260: gap of 50 bp  
50261 60819: contig of 10559 bp in length  
60820 61482: gap of 663 bp  
61483 64384: contig of 2902 bp in length  
64385 64434: gap of 50 bp  
64435 67014: contig of 2580 bp in length  
67015 67064: gap of 50 bp  
67065 93556: contig of 26492 bp in length  
93557 93606: gap of 50 bp  
93607 108482: contig of 14876 bp in length  
108483 108532: gap of 50 bp  
108533 121591: contig of 13059 bp in length  
121592 121641: gap of 50 bp  
121642 126700: contig of 5059 bp in length  
126701 126750: gap of 50 bp  
126751 128315: contig of 1565 bp in length  
128316 128365: gap of 50 bp  
128366 153628: contig of 25213 bp in length  
153629 171554: contig of 17926 bp in length  
171555 171604: gap of 50 bp  
171605 179129: contig of 7525 bp in length  
179130 179179: gap of 50 bp  
179180 182152: contig of 2973 bp in length  
182153 185199: gap of 3367 bp  
185200 188536: contig of 3017 bp in length  
188537 188974: gap of 438 bp  
188975 192508: contig of 3534 bp in length  
192509 192559: gap of 50 bp  
192560 197962: contig of 5404 bp in length  
197963 198012: gap of 50 bp  
198013 202007: contig of 3995 bp in length  
202008 202568: gap of 561 bp  
202569 211162: contig of 8594 bp in length  
211170 211212: gap of 50 bp  
211213 247082: contig of 35870 bp in length  
247083 249331: gap of 2249 bp  
249332 253480: contig of 4149 bp in length  
253481 253634: gap of 154 bp  
253635 268332: contig of 14698 bp in length  
268333 268521: gap of 189 bp  
268522 269332: contig of 1411 bp in length  
269333 270032: gap of unknown length  
270033 271073: contig of 1041 bp in length  
271074 271173: gap of unknown length  
271174 272194: contig of 1021 bp in length  
272195 272295: gap of unknown length  
272296 273957: contig of 1663 bp in length  
273958 274057: gap of unknown length  
274058 275360: contig of 1303 bp in length  
275361 275460: gap of unknown length  
275461 276754: contig of 1294 bp in length  
276755 276854: gap of unknown length  
276855 278168: contig of 1314 bp in length  
278169 278268: gap of unknown length  
278269 284106: contig of 5838 bp in length.

# FEATURES

source

Location/Qualifiers  
1. 284106  
/organism="Bos taurus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9913"  
/clone="CH240-103F13"  
3308. .3357  
/estimated\_length=50  
21883. .21932  
/estimated\_length=50  
27058. .27107  
/estimated\_length=50  
30387. .30436  
/estimated\_length=50  
34786. .34912  
/estimated\_length=127  
37141. .37190  
/estimated\_length=50  
50211. .50260  
/estimated\_length=50  
60820. .61482  
/estimated\_length=663  
64385. .64434

Query Match 100.0%; Score 17; DB 12; Length 284106;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGCCGAG 17  
|||||

Db 255390 GCCCAGCGTTGCCGAG 255406  
|||||

# RESULT 24

SC0939104/c

LOCUS SC0939104 299050 bp DNA linear BCT 16-APR-2005

DEFINITION Streptomyces coelicolor A3(2) complete genome; segment 1/29.

ACCESSION AL939104 AL109747 AL109949 AL109950 AL109962 AL109972 AL109973

AL109987 AL109989 AL110470 AL117322 AL117385 AL121854 AL590984

AL645882 AL732393

AL939104.1 GI:24413714

Streptomyces coelicolor A3(2)

Streptomyces coelicolor A3(2)

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

Streptomycineae; Streptomycetaceae; Streptomyces.

# REFERENCE

1

## AUTHORS

Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.M., Challis, G.L.,  
Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H.,  
Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M.,  
Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S.,  
Huang, C.H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neill, S.,  
Rabinowitz, E., Rajandream, M.A., Rutherford, K., Rutter, S.,  
Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S.,  
Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrell, B.G.,  
Parkhill, J. and Hopwood, D.A.

Complete genome sequence of the model actinomycete Streptomyces

coelicolor A3(2)

Nature 417 (6885), 141-147 (2002)

12000953

2 (bases 1 to 299050)

Bentley, S.D.

Direct Submission

Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces

sequencing team, Sanger Institute, Wellcome Trust Genome Campus,

Hinxton, Cambridge CB10 1SA E-mail: sdb@sanger.ac.uk

On or before Oct 26, 2002 this sequence version replaced

gi:20520931, gi:20520890, gi:20520923, gi:20520881, gi:20520932,

gi:20520933, gi:20520892, gi:20520934, gi:20520882,

gi:20520771, gi:20520935, gi:13872787, gi:20520986.

Location/Qualifiers

1. .299050

/organism="Streptomyces coelicolor A3(2)"



SLGRPELTIAVDATRSILAAVLPRYSTKAVDAALLAEMAVPHPARPTWPSALHLSR  
ASVPERLSLDERLEGAAAPVVVIVDVRGKIYLSQGFVAACEMLGYSVQAPPP  
REPOAKVVERTEGAINDLFCOHVAGHTGSNPQRGLTTAETRTWITPOLQDFLEWPI  
TCGWNQRHDLGRHVPVLFKTAITPNQMAALITSGYVFPUSGADYLELLPVRWQPI  
TERGLRDYTHVHDILDPHQRQSFVASKDGKWEVHNPHDARQIFVRLTDGQLHEI  
PMIHRDHVQFPNEAIWRHVQAEVQRGDRQHEADLADALDQLLRTRHLAETBQKT  
RRRRATRSQTAAQLPDLPCQRRPFAETAPAPAPDWSLDDLSVDVTAAGTGTSEME  
GASVPPAEAGGYGLWDAEAEQW"

## misc\_feature

8231..8623

/gene="SCO0005"

/note="Prim match to entry PF00665 rve, Integrase core  
domain, score 47.20, E-value 1.8e-12"

## gene

9304..10410

/gene="SCO0006"

/note="synonyms: SC1C9.03, SCJ30.01"

9304..10410

## CDS

Query Match 100.0%; Score 17; DB 15; Length 299050;

Best Local Similarity 100.0%; Pred. No. 2.7e+02;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACGCTTGCCGAG 17

|||||

Db 216438 GCCACGCTTGCCGAG 216422

## RESULT 25

CP000124\_29

## WPCOMMENT

Sequence split into 42 fragments LOCUS CP000124 Accession CP000124

Fragment Name Begin End

CP000124\_00 1 110000

CP000124\_01 100001 210000

CP000124\_02 200001 310000

CP000124\_03 300001 410000

CP000124\_04 400001 510000

CP000124\_05 500001 610000

CP000124\_06 600001 710000

CP000124\_07 700001 810000

CP000124\_08 800001 910000

CP000124\_09 900001 1010000

CP000124\_10 1000001 1110000

CP000124\_11 1100001 1210000

CP000124\_12 1200001 1310000

CP000124\_13 1300001 1410000

CP000124\_14 1400001 1510000

CP000124\_15 1500001 1610000

CP000124\_16 1600001 1710000

CP000124\_17 1700001 1810000

CP000124\_18 1800001 1910000

CP000124\_19 1900001 2010000

CP000124\_20 2000001 2110000

CP000124\_21 2100001 2210000

CP000124\_22 2200001 2310000

CP000124\_23 2300001 2410000

CP000124\_24 2400001 2510000

CP000124\_25 2500001 2610000

CP000124\_26 2600001 2710000

CP000124\_27 2700001 2810000

CP000124\_28 2800001 2910000

CP000124\_29 2900001 3010000

CP000124\_30 3000001 3110000

CP000124\_31 3100001 3210000

CP000124\_32 3200001 3310000

CP000124\_33 3300001 3410000

CP000124\_34 3400001 3510000

CP000124\_35 3500001 3610000

CP000124\_36 3600001 3710000

CP000124\_37 3700001 3810000

CP000124\_38 3800001 3910000

CP000124\_39 3900001 4010000

CP000124\_40 4000001 4110000

CP000124\_41 4100001 426292

Continuation (30 of 42) of CP000124 from base 2900001 (CP000124 Burkholderia pseudomallei:

Query Match 94.1%; Score 16; DB 15; Length 110000;

Best Local Similarity 100.0%; Pred. No. 9.5e+02;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CCCACGCTTGCCGAG 17

|||||

Db 43511 CCCACGCTTGCCGAG 43526

## RESULT 26

BX571965\_26

## WPCOMMENT

Sequence split into 41 fragments LOCUS BX571965 Accession BX571965

Fragment Name Begin End

BX571965\_00 1 110000

BX571965\_01 100001 210000

BX571965\_02 200001 310000

BX571965\_03 300001 410000

BX571965\_04 400001 510000

BX571965\_05 500001 610000

BX571965\_06 600001 710000

BX571965\_07 700001 810000

BX571965\_08 800001 910000

BX571965\_09 900001 1010000

BX571965\_10 1000001 1110000

BX571965\_11 1100001 1210000

BX571965\_12 1200001 1310000

BX571965\_13 1300001 1410000

BX571965\_14 1400001 1510000

BX571965\_15 1500001 1610000

BX571965\_16 1600001 1710000

BX571965\_17 1700001 1810000

BX571965\_18 1800001 1910000

BX571965\_19 1900001 2010000

BX571965\_20 2000001 2110000

BX571965\_21 2100001 2210000

BX571965\_22 2200001 2310000

BX571965\_23 2300001 2410000

BX571965\_24 2400001 2510000

BX571965\_25 2500001 2610000

BX571965\_26 2600001 2710000

BX571965\_27 2700001 2810000

BX571965\_28 2800001 2910000

BX571965\_29 2900001 3010000

BX571965\_30 3000001 3110000

BX571965\_31 3100001 3210000

BX571965\_32 3200001 3310000

BX571965\_33 3300001 3410000

BX571965\_34 3400001 3510000

BX571965\_35 3500001 3610000

BX571965\_36 3600001 3710000

BX571965\_37 3700001 3810000

BX571965\_38 3800001 3910000

BX571965\_39 3900001 4010000

BX571965\_40 4000001 4074542

## Query Match

94.1%; Score 16; DB 15; Length 110000;

Best Local Similarity 100.0%; Pred. No. 9.5e+02;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CCCACGCTTGCCGAG 17

|||||

Db 73449 CCCACGCTTGCCGAG 73464

## RESULT 27

CP000010\_16

## WPCOMMENT

Sequence split into 36 fragments LOCUS CP000010 Accession CP000010

Fragment Name Begin End

CP000010\_00 1 110000  
CP000010\_01 100001 210000  
CP000010\_02 200001 310000  
CP000010\_03 300001 410000  
CP000010\_04 400001 510000  
CP000010\_05 500001 610000  
CP000010\_06 600001 710000  
CP000010\_07 700001 810000  
CP000010\_08 800001 910000  
CP000010\_09 900001 1010000  
CP000010\_10 1000001 1110000  
CP000010\_11 1100001 1210000  
CP000010\_12 1200001 1310000  
CP000010\_13 1300001 1410000  
CP000010\_14 1400001 1510000  
CP000010\_15 1500001 1610000  
CP000010\_16 1600001 1710000  
CP000010\_17 1700001 1810000  
CP000010\_18 1800001 1910000  
CP000010\_19 1900001 2010000  
CP000010\_20 2000001 2110000  
CP000010\_21 2100001 2210000  
CP000010\_22 2200001 2310000  
CP000010\_23 2300001 2410000  
CP000010\_24 2400001 2510000  
CP000010\_25 2500001 2610000  
CP000010\_26 2600001 2710000  
CP000010\_27 2700001 2810000  
CP000010\_28 2800001 2910000  
CP000010\_29 2900001 3010000  
CP000010\_30 3000001 3110000  
CP000010\_31 3100001 3210000  
CP000010\_32 3200001 3310000  
CP000010\_33 3300001 3410000  
CP000010\_34 3400001 3510000  
CP000010\_35 3500001 3510148

Continuation (17 of 36) of CP000010 from base 1600001 (CP000010 Burkholderia mallei ATCC

Query Match 94.1%; Score 16; DB 15; Length 110000;

Best Local Similarity 100.0%; Pred. No. 9.5e+02;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 CCCACGTTGCCGAG 17

Db 98066 CCCACGTTGCCGAG 98081

RESULT 28

AC167299/c

LOCUS AC167299 174298 bp DNA linear HTG 22-OCT-2005

DEFINITION Oryctolagus cuniculus clone LB1-35017, WORKING DRAFT SEQUENCE, 6

ordered pieces.

AC167299

VERSION AC167299.3 GI:78042356

KEYWORDS HTG; HTGS\_PHASE2; HTGS\_DRAFT.

SOURCE Oryctolagus cuniculus (rabbit)

ORGANISM Oryctolagus cuniculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Lagomorpha;

Leporidae; Oryctolagus.

1 (bases 1 to 174298).

Antoneillis,A., Ayete,K., Bass,D., Benjamin,B., Bera,J.,

Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S., Chu,G.,

Coleman,H., Franks,S., Fuksenko,T., Gestole,M., Greene,A., Guan,X.,

Gupta,J., Gurson,N., Haghighi,P., Han,E., Han,J., Hansen,N.,

Ho,S.-I., Hu,P., Hunter,G., Hurle,B., Idol,J.R., Kwong,P.,

Laric,P., Larson,S., Lee-lin,S.-Q., Legaspi,R., Madden,M.,

Maduro,Q.D., Maduro,V.B., Margulies,E.H., Mastello,C., Maskeri,B.,

McDowell,J., Montemayor,C., Mullikin,J.C., Park,M., Portnoy,M.E.,

Prasad,A., Puri,O., Rantz,K., Reddix-Dugue,N., Sante,A.,

Schandler,K., Schueler,M.G., Sison,C., Stantripop,S., Teye,A.,

Thomas,J.W., Thomas,P.J., Tsipouri,V., Ung,L., Vogt,J.L.,

Wetherby,K.D., Withers,T.R., Young,A. and Green,E.D.

NISC Comparative Sequencing Initiative

Unpublished

2 (bases 1 to 174298)

Green,E.D.

Direct Submission

Submitted (24-AUG-2005) NIH Intramural Sequencing Center, 5625

Fishers Lane, Rockville, MD 20852, USA

3 (bases 1 to 174298)

Green,E.D.

Direct Submission

Submitted (22-OCT-2005) NIH Intramural Sequencing Center, 5625

Fishers Lane, Rockville, MD 20852, USA

On Oct 22, 2005 this sequence version replaced gi:77627864.

----- Genome Center

Center: NIH Intramural Sequencing Center

Center code: NISC

Web site: http://www.nisc.nih.gov

Contact: nisc\_zoo@nhgri.nih.gov

----- Project Information

Center project name: lnj

Center clone name: 035017

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is generally based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 172944 bases at least Q40

Consensus quality: 173403 bases at least Q30

Consensus quality: 173643 bases at least Q20

Insert size: 176000; agarose-fp

Insert size: 173798; sum-of-contigs

Quality coverage: 9.18x in Q20 bases; agarose-fp

Quality coverage: 9.29x in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 6 contigs. Gaps between the contigs

\* are represented as runs of N. The order of the pieces

\* is believed to be correct as given, however the sizes

\* of the gaps between them are based on estimates that have

\* provided by the submitter.

\* This sequence will be replaced

\* by the finished sequence as soon as it is available and

\* the accession number will be preserved.

\* 1 56412: contig of 56412 bp in length

\* 56413 56512: gap of unknown length

\* 56513 61819: contig of 5307 bp in length

\* 61820 61919: gap of unknown length

\* 61920 68051: contig of 6132 bp in length

\* 68052 68151: gap of unknown length

\* 68152 101952: contig of 33801 bp in length

\* 101953 102052: gap of unknown length

\* 102053 157436: contig of 55384 bp in length

\* 157437 157536: gap of unknown length

\* 157537 174298: contig of 16762 bp in length.

Location/Qualifiers

1..174298

/organism="Oryctolagus cuniculus"

/mol\_type="genomic DNA"

/db\_xref="taxon:9986"

/clones="LB1-35017"

/clone\_lib="LB1"

FEATURES  
Source



/note="BAC resource: http://bacpac.chori.org/  
breed: New Zealand White"

## misc\_feature

1..56412  
/note="assembly\_fragment  
clone\_end:177  
vector\_side:left"

## gap

56413..56512  
/estimated\_length=unknown

## misc\_feature

56513..61819  
/note="assembly\_fragment"

## gap

61820..61919  
/estimated\_length=unknown

## misc\_feature

61920..68051  
/note="assembly\_fragment"

## gap

68052..68151  
/estimated\_length=unknown

## misc\_feature

68152..101952  
/note="assembly\_fragment"

## gap

101953..102052  
/estimated\_length=unknown

## misc\_feature

102053..157436  
/note="assembly\_fragment"

## gap

157437..157536  
/estimated\_length=unknown

## misc\_feature

157537..174298  
/note="assembly\_fragment  
vector\_side:right"

## ORIGIN

Query Match 94.1%; Score 16; DB 12; Length 174298;  
Best Local Similarity 100.0%; Pred. No. 1e+03;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGA 16

Db 87377 GCCCAGCGTTGGCCGA 87362

## RESULT 29

MAU564400/c MAU564400 117 bp DNA linear BCT 15-APR-2005  
LOCUS Mycobacterium aurum partial GyrB gene for DNA gyrase subunit B,  
DEFINITION strain ATCC 23366.

ACCESSION AJ564400.1 GI:31559458

VERSION DNA gyrase subunit B; GyrB gene.

KEYWORDS Mycobacterium aurum

SOURCE Mycobacterium aurum

ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

REFERENCE 1 Corynebacterineae; Mycobacteriaceae; Mycobacterium.

AUTHORS Dauendorffer, J.N., Guillemin, I., Aubry, A., Truffot-Pernot, C.,

Sougakoff, W., Jarlier, V., and Cambau, E.

TITLE Identification of mycobacterial species by PCR sequencing of

quinolone resistance-determining regions of DNA gyrase genes

J. Clin. Microbiol. 41 (3), 1311-1315 (2003)

PUBMED 12624075

REFERENCE 2 (bases 1 to 117)

AUTHORS Cambau, E.

TITLE Direct Submission

Submitted (26-FEB-2003) Cambau E., Bacteriologie, Faculte de

Medecine Pitie-Salpetriere, 91 Boulevard de l'Hopital, 75013 Paris,

FRANCE

## FEATURES

Location/Qualifiers

1..117

/organism="Mycobacterium aurum"

/mol\_type="genomic DNA"

/strain="ATCC 23366"

/db\_xref="taxon:1791"

<1..>117

/gene="GyrB"

<1..>117

## gene

## CDS

/gene="GyrB"  
/codon\_start=1  
/transl\_table=11  
/product="DNA gyrase subunit B"  
/protein\_id="CAD92772.1"  
/db\_xref="GI:31559458"  
/db\_xref="UniProtKB/TREMBL:Q7WTA2"  
/translation="DSAGGSAKSGRDSFQAILPLRGKIINVEKARIDRVLKN"

## ORIGIN

Query Match 90.6%; Score 15.4; DB 15; Length 117;  
Best Local Similarity 94.1%; Pred. No. 9.4e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17

Db 31 GCCCAGCTTGGCCGAG 15

## RESULT 30

MAU564400/c MAU564400 385 bp DNA linear PRI 11-JAN-2006  
LOCUS Mycobacterium aurum partial HLA-D gene for MHC class II antigen,  
DEFINITION HLA-DQB1\*06new allele, exon 2.

ACCESSION AM181332

VERSION AM181332.1 GI:84794315

KEYWORDS HLA-D gene; HLA-DQB1\*06new allele; human leucocyte antigen D; major

histocompatibility complex; MHC class II antigen.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.

REFERENCE 1 Witter, K.

A novel HLA-DQB1 allele, DQB1\*06xx was detected through routine HLA

class II high resolution typing

Unpublished

Witter, K.

Direct Submission

Submitted (10-JAN-2006) Witter K., Labor of Immunogenetics,

University Munich, Paul-Heyse Strasse 33, 80336 Munich, GERMANY

## FEATURES

Location/Qualifiers

1..385

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/isolate="158308"

/db\_xref="taxon:9606"

<1..>283

/gene="HLA-D"

<1..>283

/allele="HLA-D"

/allele="HLA-DQB1\*06new"

<1..>283

/gene="HLA-D"

/function="antigen presentation"

/function="antigen presenting molecule"

/codon\_start=3

/allele="HLA-DQB1\*06new"

/product="MHC class II antigen"

/protein\_id="CAJ57391.1"

/db\_xref="GI:84794316"

/translation="DFVQFGKMGCFYNGTERVRLVTRHIYNREYVRFSDVGVYRA

VTQGRPDAEYWNYSQKEVLEGTAEELDTVCRHNYEVAFRGILQRRGERR"

1..283

/gene="HLA-D"

/number=2

/allele="HLA-DQB1\*06new"

## Query Match

Best Local Similarity 90.6%; Score 15.4; DB 5; Length 385;

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCCGAG 15

Db 31 GCCCAGCTTGGCCGAG 15

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17  
 ||||| ||||| |||||  
 Db 296 GCCCAGCGTTGGCCGAG 312

RESULT 31  
 AY535046/c  
 LOCUS AY535046 586 bp DNA linear ENV 01-MAR-2004  
 DEFINITION Uncultured eukaryote clone E8 18S ribosomal RNA gene, partial  
 sequence.  
 ACCESSION AY535046  
 VERSION AY535046.1 GI:42794960  
 KEYWORDS ENV.  
 SOURCE uncultured eukaryote  
 ORGANISM uncultured eukaryote  
 Eukaryota; environmental samples.  
 REFERENCE 1 (bases 1 to 586)  
 AUTHORS Duplessis,M.R., Dufour,S.C., Blankenship,L.E., Yayanos,A.A. and Felbeck,H.  
 CONSRTM SIO  
 TITLE Anatomical and Experimental Evidence for Particulate Feeding in Lucinoma aequizonata and Parvilucina tenuisculpta (Bivalvia: Lucinidae) from the Santa Barbara Basin  
 JOURNAL Marine Biology (2004) In press  
 REFERENCE 2 (bases 1 to 586)  
 AUTHORS Blankenship,L.E., Duplessis,M.R., Dufour,S.C., Yayanos,A.Aristides. and Felbeck,H.  
 CONSRTM SIO  
 TITLE Direct Submission  
 JOURNAL Submitted (29-JAN-2004) Marine Biology Research Division, Scripps Institution of Oceanography, 9500 Gilman Dr. 0202, La Jolla, CA 92093, USA

FEATURES  
 source  
 1..586  
 /organism="uncultured eukaryote"  
 /mol\_type="genomic DNA"  
 /isolation\_source="Lucinoma aequizonata gut"  
 /db\_xref="taxon:100272"  
 /clones="E8"  
 /environmental sample  
 complement(<1..586)  
 /product="18S ribosomal RNA"

rRNA

ORIGIN

Query Match 90.6%; Score 15.4; DB 1; Length 586;  
 Best Local Similarity 94.1%; Pred. No. 1.1e+03;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17  
 ||||| ||||| |||||  
 Db 283 GCCCAGCGTTGGCCGAG 267

RESULT 32  
 DQ288979/c  
 LOCUS DQ288979 738 bp DNA linear VRL 05-DEC-2005  
 DEFINITION Bovine leukemia virus strain 991 integrase-like (pol) gene, partial  
 sequence.  
 ACCESSION DQ288979  
 VERSION DQ288979.1 GI:82791298  
 KEYWORDS  
 SOURCE Bovine leukemia virus  
 ORGANISM Bovine leukemia virus  
 Viruses; Retro-transcribing viruses; Retroviridae;  
 Orthoretrovirinae; Deltaretrovirus.  
 REFERENCE 1 (bases 1 to 738)  
 AUTHORS Hirsch,C., Barbosa-Stancioli,E.F., Camargos,M.F., Reis,J.K.P. and Leite,R.C.  
 TITLE Bovine leukemia virus pol gene: Genetic variability and phylogeny of the integrase region of Brazilian samples  
 JOURNAL Unpublished

REFERENCE 2 (bases 1 to 738)  
 AUTHORS Hirsch,C., Barbosa-Stancioli,E.F., Camargos,M.F., Reis,J.K.P. and Leite,R.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (14-NOV-2005) Medicina Veterinaria Preventiva, Escola de Veterinaria da Universidade Federal de Minas Gerais, Avenida Antonio Carlos 6627. Bairro Sao Francisco caixa postal 567., Belo Horizonte, Minas Gerais 30.123-970, Brasil

FEATURES  
 Location/Qualifiers  
 source  
 1..738  
 /organism="Bovine leukemia virus"  
 /proviral  
 /mol\_type="genomic DNA"  
 /strain="991"  
 /specific\_host="Bos taurus"  
 /db\_xref="taxon:11901"  
 /country="Brazil: Pedro Leopoldo, Minas Gerais State"  
 <1..>738  
 /gene="pol"  
 misc\_feature  
 <1..>738  
 /gene="pol"  
 /note="similar to integrase"

ORIGIN

Query Match 90.6%; Score 15.4; DB 10; Length 738;  
 Best Local Similarity 94.1%; Pred. No. 1.2e+03;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17  
 ||||| ||||| |||||  
 Db 188 GCCCAGCGTTGGCCGAG 172

RESULT 33  
 AY381191  
 LOCUS AY381191 1802 bp DNA linear ENV 08-JUN-2004  
 DEFINITION Uncultured eukaryote clone BL000921.11 18S rRNA gene, complete  
 sequence.  
 ACCESSION AY381191  
 VERSION AY381191.1 GI:39981844  
 KEYWORDS ENV.  
 SOURCE uncultured marine eukaryote  
 ORGANISM uncultured marine eukaryote  
 Eukaryota; environmental samples.  
 REFERENCE 1 (bases 1 to 1802)  
 AUTHORS Massana,R., Castresana,J., Balague,V., Guillou,L., Romari,K., Groisillier,A., Valentin,K. and Pedros-Allo,C.  
 TITLE Phylogenetic and Ecological Analysis of Novel Marine Stramenopiles Appl. Environ. Microbiol. 70 (6), 3528-3534 (2004)  
 JOURNAL 15184153  
 PUBMED  
 REFERENCE 2 (bases 1 to 1802)  
 AUTHORS Massana,R., Castresana,J., Balague,V., Guillou,L., Romari,K., Groisillier,A., Valentin,K. and Pedros-Allo,C.  
 TITLE Direct Submission  
 JOURNAL Submitted (03-SEP-2003) Marine Biology and Oceanography, Institut de Ciències del Mar, CSIC, Passeig Marítim de la Barceloneta 37-49, Barcelona, Catalonia 08003, Spain

FEATURES  
 Location/Qualifiers  
 source  
 1..1802  
 /organism="uncultured marine eukaryote"  
 /mol\_type="genomic DNA"  
 /isolation\_source="coastal surface water, Northwestern Mediterranean, Bienes Bay"  
 /db\_xref="taxon:203449"  
 /clones="BL000921.11"  
 /environmental\_sample  
 1..1802  
 /product="18S ribosomal RNA"

rRNA

ORIGIN

Query Match 90.6%; Score 15.4; DB 1; Length 1802;  
 Best Local Similarity 94.1%; Pred. No. 1.3e+03;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	1	GCCACGCTTGGCCGAG	17
Db	1496	GCACACGCTTGGCCGAG	1512
RESULT 34			
AX416413/c			
LOCUS	AX416413	2306 bp	DNA linear PAT 02-SEP-2002
DEFINITION	Sequence 3404 from Patent WO0228891.		
ACCESSION	AX416413		
VERSION	AX416413.1	GI:21448870	
KEYWORDS			
SOURCE	Listeria monocytogenes ATCC 19115		
ORGANISM	Listeria monocytogenes ATCC 19115		
REFERENCE	1	Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.	
AUTHORS	Kunst, F. and Glaeser, P.		
TITLE	Listeria innocua, genome and applications		
JOURNAL	Patent: WO 0228891-A 3404 11-APR-2002;		
FEATURES	SCIENTIFIQUE (CNRS) (FR)		
source	Location/Qualifiers		
1..2306	/organism="Listeria monocytogenes ATCC 19115"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:176281"		
ORIGIN			
Query Match	90.6%; Score 15.4; DB 2; Length 2306;		
Best Local Similarity	94.1%; Pred. No. 1.4e+03;		
Matches	16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;		
Qy	1	GCCACGCTTGGCCGAG	17
Db	834	GCCACGCTTGGCTGAG	818
RESULT 35			
HSA564991			
LOCUS	HSA564991	2550 bp	DNA linear PRI 27-JUL-2005
DEFINITION	Homo sapiens HLA-DQB1 gene for MHC class II antigen, HLA-DQB1*0602 allele, intron 2.		
ACCESSION	AJ564991		
VERSION	AJ564991.1	GI:3145385	
KEYWORDS	HLA-DQB1 gene; HLA-DQB1*0602 allele; major histocompatibility complex; MHC class II antigen.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE	1	Dunn, P.P., Day, S., Williams, S. and Bendukidze, N.	
AUTHORS	HLA-DQB1 sequencing-based typing using newly identified conserved nucleotide sequences in introns 1 and 2		
TITLE	Tissue Antigens 66 (2), 99-106 (2005)		
JOURNAL	PUBMED		
REFERENCE	16029429		
AUTHORS	2 (bases 1 to 2550)		
TITLE	Dunn, P.P.		
JOURNAL	Direct Submission		
FEATURES	Submitted (03-JUN-2003) Dunn P.P., H & I DNA Reference Laboratory, National Blood Service, Southmead Road, BS10 5ND, UNITED KINGDOM		
source	Location/Qualifiers		
1..2550	/organism="Homo sapiens"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:9606"		
	/cell_line="W6V"		
	/gene="HLA-DQB1"		
	1..2550		
	/gene="HLA-DQB1"		
	1..2550		
	/function="antigen presenting molecule"		
	/allele="HLA-DQB1*0603"		
ORIGIN			
Query Match	90.6%; Score 15.4; DB 5; Length 2550;		
Best Local Similarity	94.1%; Pred. No. 1.4e+03;		
Matches	16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;		
Qy	1	GCCACGCTTGGCCGAG	17
Db	26	GCCACCCCTTGCCGAG	42
RESULT 37			
HSA564993			
LOCUS	HSA564993	2550 bp	DNA linear PRI 27-JUL-2005
DEFINITION	Homo sapiens HLA-DQB1 gene for MHC class II antigen, HLA-DQB1*060401 allele, intron 2.		
ACCESSION	AJ564993		
VERSION	AJ564993.1	GI:31455387	
KEYWORDS	HLA-DQB1 gene; HLA-DQB1*060401 allele; major histocompatibility complex; MHC class II antigen.		
SOURCE	Homo sapiens (human)		
ORIGIN			
Query Match	90.6%; Score 15.4; DB 5; Length 2550;		
Best Local Similarity	94.1%; Pred. No. 1.4e+03;		
Matches	16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;		
Qy	1	GCCACGCTTGGCCGAG	17
Db	26	GCCACCCCTTGCCGAG	42
RESULT 39			
HSA564993			
LOCUS	HSA564993	2550 bp	DNA linear PRI 27-JUL-2005
DEFINITION	Homo sapiens HLA-DQB1 gene for MHC class II antigen, HLA-DQB1*060401 allele, intron 2.		
ACCESSION	AJ564993		
VERSION	AJ564993.1	GI:31455387	
KEYWORDS	HLA-DQB1 gene; HLA-DQB1*060401 allele; major histocompatibility complex; MHC class II antigen.		
SOURCE	Homo sapiens (human)		
ORIGIN			
Query Match	90.6%; Score 15.4; DB 5; Length 2550;		
Best Local Similarity	94.1%; Pred. No. 1.4e+03;		
Matches	16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;		
Qy	1	GCCACGCTTGGCCGAG	17
Db	26	GCCACCCCTTGCCGAG	42
RESULT 41			
HSA564991			
LOCUS	HSA564991	2550 bp	DNA linear PRI 27-JUL-2005
DEFINITION	Homo sapiens HLA-DQB1 gene for MHC class II antigen, HLA-DQB1*0602 allele, intron 2.		
ACCESSION	AJ564991		
VERSION	AJ564991.1	GI:3145385	
KEYWORDS	HLA-DQB1 gene; HLA-DQB1*0602 allele; major histocompatibility complex; MHC class II antigen.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE	1	Dunn, P.P., Day, S., Williams, S. and Bendukidze, N.	
AUTHORS	HLA-DQB1 sequencing-based typing using newly identified conserved nucleotide sequences in introns 1 and 2		
TITLE	Tissue Antigens 66 (2), 99-106 (2005)		
JOURNAL	PUBMED		
REFERENCE	16029429		
AUTHORS	2 (bases 1 to 2550)		
TITLE	Dunn, P.P.		
JOURNAL	Direct Submission		
FEATURES	Submitted		

ORGANISM	Homo sapiens	JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.	PUBMED	12477932
AUTHORS	1 Dunn, P.P., Day, S., Williams, S. and Bendukidze, N.	REFERENCE	2 (bases 1 to 2623)
TITLE	HLA-DQB1 sequencing-based typing using newly identified conserved nucleotide sequences in introns 1 and 2	AUTHORS	Strausberg, R.
JOURNAL	Tissue Antigens 66 (2), 99-106 (2005)	TITLE	Direct Submission
PUBMED	16029429	JOURNAL	Submitted (07-DEC-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REFERENCE	2 (bases 1 to 2550)	REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov
AUTHORS	Dunn, P.P.	COMMENT	On Dec 9, 2003 this sequence version replaced gi:17511964.
TITLE	Direct Submission	Contact:	MGC help desk
JOURNAL	Submitted (03-JUN-2003) Dunn P.P., H & I DNA Reference Laboratory, National Blood Service, Southmead Road, BS10 5ND, UNITED KINGDOM	Email:	cgabs@femail.nih.gov
FEATURES	Location/Qualifiers	Tissue Procurement:	ATCC
source	1..2550	cDNA Library Preparation:	Rubin Laboratory
	/organism="Homo sapiens"	DNA Sequencing by:	The I.M.A.G.E. Consortium (LLNL)
	/mol_type="genomic DNA"	Sequencing Center (NISC),	
	/db_xref="taxon:9606"	Gaithersburg, Maryland;	
gene	1..2550	Web site:	http://www.nisc.nih.gov/
intron	1..2550	Contact:	nisc_mgc@nhgri.nih.gov
	/gene="HLA-DQB1"	Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,	
	/number=2	Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooke, S.,	
	/functions="antigen presenting molecule"	Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,	
	/allele="HLA-DQB1*060401"	Hansen, N., Ho, S.-L., Karling, E., Kwong, P., Laric, P., Legaspi, R.,	
ORIGIN		Maduro, Q.L., Masiello, C., Maskeri, B., Mastriopop, S., Thomas, P.J., Touchman, J.W.,	
		McDowell, J., Pearson, R., Stantropop, S., Thomas, P.J., Touchman, J.W.,	
		Tsurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,	
		Young, A., Zhang, L.-H. and Green, E.D.	
Query Match	90.6%; Score 15.4; DB 5; Length 2550;	Clone distribution:	MGC clone distribution information can be found
Best Local Similarity	94.1%; Pred. No. 1.4e+03;	through the I.M.A.G.E. Consortium/LLNL at:	http://image.llnl.gov
Matches	16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	Series:	IRAL Plate: 20 Row: f Column: 6.
Qy	1 GCCCAGCTTGCCGCGAG 17	Location/Qualifiers	
Db	26 GCCCAGCTTGCCGCGAG 42	1..2623	
RESULT 38		/organism="Homo sapiens"	
BC018939		/mol_type="mRNA"	
LOCUS	BC018939 2623 bp mRNA linear PRI 24-FEB-2004	/db_xref="taxon:9606"	
DEFINITION	Homo sapiens cartilage intermediate layer protein 2, mRNA (CDNA clone IMAGE:4126917), partial cds.	/clones="IMAGE:4126917"	
ACCESSION	BC018939	/tissue_type="Brain, neuroblastoma"	
VERSION	BC018939.2 GI:39644584	/clone_id="NIH MGC 19"	
KEYWORDS		/lab_host="DH10B-R"	
SOURCE	Homo sapiens (human)	/note="Vector: pOTB7"	
ORGANISM		<1..2623	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.		/gene="CILP2"	
1 (bases 1 to 2623)		/note="synonyms: MGC45771, CLIP-2"	
Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,		/db_xref="GeneID:148113"	
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,		<1..1964	
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,		/gene="CILP2"	
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,		/codon_start=3	
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,		/product="CILP2 protein"	
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,		/protein_id="AAH18939.2"	
Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,		/db_xref="GI:39644585"	
Carninci, P., Prange, C., Raha, S.S., Igoe, N.A., Peters, G.J.,		/translation="AYOGDSTIEVPPSTQRLVTVFDPSPGFMDAVRLPDPGRGAV	
Abrahamson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,		YHEVTKKAPVILHTSQSNTLPLEGDEAPLCELVPSGARFRADKCYSGPVEA	
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,		RVTFVDPDLTSAASAPDLRFVDSDELAPLRTYGMFSDLRAPGSAEQVQGPVAV	
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,		RVAASQIHMFGHVEALKLSINPETLWEEESGFRREGSGRVRRERVLGVNGVEI	
Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,		RERLFLNDVPRRCFCVKYAVANDKFTSEQVEGVVTVLNLPEAPGFSANPRAW	
Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,		RFDASVTGNGACLPACDADRDPAVLTALVTATLGGELPAPSLPRPLPATVGTQV	
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,		YHFRPVEADKYENVVFRFGTASWTGDLAWPFPQEPFACELKVKIQGPQEVYV	
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,		RSHNAGSHPRTRGQLYGLRDARSVPDRPFRTSAACVEFKCSGMLFPQQRVRLVT	
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalilus, D.E.,		IMPOGSCRRVAVNGLRLDYLTRHPFPVPAEDPFAAFSLAPLDPLGHNGVTVTDQSP	
Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.		RLAKEIAIGRCFDGSDGFSREMKADAGTAVTFCREPPAGRPRLFORLLESPTALG	
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences		DIREMSEAAQAQASGSLRTRGRVRVQ"	
TITLE		ORIGIN	
		Query Match	90.6%; Score 15.4; DB 5; Length 2623;
		Best Local Similarity	94.1%; Pred. No. 1.4e+03;
		Matches	16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
		Qy	1 GCCCAGCTTGCCGCGAG 17

```
Db      1696 GCCCAGCTTGGCCAAG 1712
|||||
BC034926      2623 bp      mRNA      linear      PRI 27-APR-2005
Homo sapiens cartilage intermediate layer protein 2, mRNA (CDNA
clone IMAGE:4124496), partial cds.
BC034926
BC034926.1 GI:22028437
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
1 (bases 1 to 2623)
Straussberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,P.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Ditachenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Armstrong,R.D., Mulhaly,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettner,M., Madan,A., Young,A.C., Rodrigues,S.,
Sanchez,A., Whitting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzyzanski,M.I., Skalska,U., Smallus,D.E.,
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Mammalian Gene Collection Program Team
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 2623)
NIH MGC Project
Direct Submission
Submitted (29-JUL-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@bcgsc.bc.ca
Martin Hirst, Thomas Zeng, Ryan Morin, Michelle Moksa, Johnson
Pang, Diana Mah, Jing Wang, Kieth Fichter, Eric Chuah, Allen
Delaney, Rob Kirkpatrick, Agnes Baross, Sarah Barber, Mabel
Brown-John, Steve S. Chand, William Chow, Ryan Babakaiff, Dave
Wong, Corey Matsuo, Jaclyn Belland, Susan Gibson, Luis delRio, Ruth
Featherstone, Malachi Griffith, Obi Griffith, Ran Guin, Nancy Liao,
Kim MacDonald, Mike R. Mayo, Josh Moran, Diana Palmquist, JR
Santos, Duane Smalish, Jeff Stott, Miranda Tsai, George Yang,
Jacquie Schein, Asim Siddiqui, Steven Jones, Rob Holt, Marco Marra.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IPAL Plate: 43 Row: d Column: 23.
Location/Qualifiers
1. .2623
/organism="Homo sapiens"
/mol_type="mRNA"
FEATURES
source
db_xref="taxon:9606"
/clone="IMAGE:4124496"
/tissue_type="Brain, neuroblastoma"
/clone_lib="NIH_MGC_19"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
<1. .2623
/gene="CILP2"
/note="synonyms: CLIP-2, MGC45771"
/db_xref="GeneID:148113"
<1. .1964
/gene="CILP2"
/codon_start=3
/product="CILP2 protein"
/protein_id="AAH34926.2"
/db_xref="GI:62914016"
/db_xref="GeneID:148113"
/translat="AYOGDFTIEVPPSTQRLVVTFFVDPGSEFMDAVRVLFPDPRGAGV
YHEVKMRKAPVILHTSQSTIPLGEDEAPLCELVLPSCAFRRADGKPYSGPVEA
RVTVDPDLTSAASAPSLRFDVSDGELAPLRTYGMFSVDLRAFGSAEQLOQGVAV
RVAASQIHMPGHVEALKLNSLPETGLWEESGFRREGSGRRVRERERVELVGNVEI
RERLENDLVPERRRCFVKRAYANDKFTPSEQVEGVVTVLNLDEPAPGFSANPRAWG
RFDGATGPNAGCLPACDADRDPDAYTALTATLGTGGELEPAPSLRPLPATVGTQP
YLDRLGYRTDHPDKFRNGFRINLAKPRDGLPAEANGPVYWRSLRCCQCAPVTAS
HFRARVEADKYENVVYFPRGTPTASMTGDLAWNPQEFACFLKVKIOGPQBYMV
RSHNAGSHPTRGOLYLRDARSVRDPERPDTSAACVEFKCSGMLPQORVDRLTVP
IMPQSCRRVAVNGLLRDYLRPHPPPAEDPAAFSLAPLDPLGHNYGVYTVTDQSP
RLAKEIAIGRCFDGSSGFSREMKADAGTAVTFOCRPEPPAGRPSPFLQRLLESPTALG
DIRMESEAAQAASGSLTRRRVRQ"
ORIGIN
Query Match 90.6%; Score 15.4; DB 5; Length 2623;
Best Local Similarity 94.1%; Pred. No. 1.4e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GCCCAGCTTGGCCGAG 17
|||||
Db 1696 GCCCAGCTTGGCCAAG 1712
RESULT 40
AF420019
LOCUS AF420019 3252 bp DNA linear PLN 09-JAN-2003
DEFINITION Aspergillus nidulans 1,4-beta-D-glucan-cellobiohydrolase (cbha)
gene, complete cds.
ACCESSION AF420019.1 GI:21449326
VERSION AF420019.1
KEYWORDS
SOURCE
ORGANISM
Emericella nidulans (anamorph: Aspergillus nidulans)
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; Emericella.
1 (bases 1 to 3252)
Lockington,R.A., Roddourn,L., Barnett,S., Carter,C.J. and
Kelly,J.M.
Regulation by carbon and nitrogen sources of a family of cellulases
in Aspergillus nidulans
Fungal Genet. Biol. 37 (2), 190-196 (2002)
12409103
2 (bases 1 to 3252)
Lockington,R.A., Roddourn,L., Barnett,S., Carter,C.C. and
Kelly,J.M.
Direct Submission
Submitted (18-SEP-2001) Molecular Biosciences, University of
Adelaide, North Terrace, Adelaide, SA 5005, Australia
Location/Qualifiers
1. .3252
/organism="Emericella nidulans"
/mol_type="genomic DNA"
/db_xref="taxon:162425"
/note="anamorph: Aspergillus nidulans"
1231. .2884
/gene="cbha"
gene
```

```
mRNA
join(1231..1498,1555..1938,1995..2884)
/gene="cbha"
/product="1,4-beta-D-glucan-cellobiohydrolase"
join(1283..1498,1555..1938,1995..2735)
/gene="cbha"
/EC_number="3.2.1.91"
/notes="cellulase"
/codon_start=1
/product="1,4-beta-D-glucan-cellobiohydrolase"
/protein_id="AAMS4069.1"
/db_xref="GI:21449327"
/translation="MYQALLFSALLSVRAQAQAGTAQEVHPSLTWQCEASGSCTE
VAGSVLDSNWRWTHSVGTYNCTYGNEDWATLCPDNESCQACAVDGADEYATYGIT
SNGDSLTKVTSNVSGRVYLMEDDETQYMFLLNNEFTFDVDSNFPCLGALAYF
TSMADGGLSKYEAGTAKYGTGCDQCPCQPRDIKFKINGLGNVEGPEPSDANAGVG
GMGTCCPEMDIWEANSISTATYTHPCDSVEQTCWCEGSCGTYSDDRYGCTCPDGDG
PNSYRWGNTFYPGAIIDTSSKFTVVTQFIADGGSLSSEIKREYVONGEVI PNSENI
SGVEGNSITSEFTAQKTARFDEIFAQHGLSAMGDAAAMVLLISINDDHHSSMMW
LDSSYPTDADPSQPGVARGTCQAGQPDVVESEHADASVTFNSNIKEFGPIGSTF"

ORIGIN
Query Match          90.6%; Score 15.4; DB 4; Length 3252;
Best Local Similarity 94.1%; Pred. No. 1.4e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCACGCTTGGCCGAG 17
||||| |||||||
Db 1207 GCCCACGCTTGGCCGAG 1223

RESULT 41
CQ727552          3325 bp      DNA      linear      PAT 03-FEB-2004
LOCUS
DEFINITION      Sequence 13486 from Patent WO02068579.
ACCESSION      CQ727552
VERSION      CQ727552.1 GI:42293852
KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.

REFERENCE
AUTHORS      Venter,C.J., Adams,M.C., Li,P.W. and Myers,E.W.
TITLE      Kits, such as nucleic acid arrays, comprising a majority of
humanexons or transcripts, for detecting expression and other uses
thereof
JOURNAL      Patent: WO 02068579-A 13486 06-SEP-2002;
PE Corporation Location/Qualifiers
FEATURES
source
1..3325
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match          90.6%; Score 15.4; DB 2; Length 3325;
Best Local Similarity 94.1%; Pred. No. 1.4e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCACGCTTGGCCGAG 17
||||| |||||||
Db 3095 GCCCACGCTTGGCCGAG 3111

RESULT 42
AF542080          3602 bp      mRNA      linear      PRI 25-SEP-2002
LOCUS
DEFINITION      Homo sapiens cartilage intermediate layer protein-like protein
CUIP-2 mRNA, complete cds.
ACCESSION      AF542080
VERSION      AF542080.1 GI:23307794
KEYWORDS
```

```
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.

REFERENCE
AUTHORS      Crowl,R.M., Szeto,D. and Farley,D.
TITLE      CILP-2, an isoform of cartilage intermediate layer protein
JOURNAL      Unpublished
REFERENCE
AUTHORS      Crowl,R.M., Szeto,D. and Farley,D.
TITLE      Direct Submission
JOURNAL      Submitted (29-AUG-2002) Arthritis Research, Novartis
Pharmaceuticals, 556 Morris Ave, Summit, NJ 07901, USA

FEATURES
source
1..3602
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="19"
/map="19p13.11"
/tissue_type="osteochondritic cartilage"
54..3524
/notes="glycoprotein; related to cartilage matrix protein
CILP"
/codon_start=1
/product="cartilage intermediate layer protein-like
protein CILP-2"
/protein_id="AANI17826.1"
/db_xref="GI:23307795"
/translation="MASLLPLGLCVVAHLAGADATPTTEPMATALGLERRSVYTC
QPSPALDEEASESWFNVDHPGGDGFESLAAIRFYVGPARKVCRPLALERTTD
WALPSAVGERVHLNPTRGFWCLNREQPRRCNSYHVRFCPLEASGAWGPGPCSG
SCGPFRLRRRHCHPCSPAGDACPGRLPAQKCVRCPCGSLDTCEDPHLLGSVVTP
SGOPLLGARVSLDQPGTAVTSDAHGTFRVPGCADSRANIRAOQMGDFSAGEAQAQAN
GSI SVVTIILDKLEKPYLVKHPSRVREAGQNTVFCCKASGTPMKVYSWFHNGTLLD
RRAHGVAHLELRLGRPDQAGIYHCKAWNAGAVRSCTARTLVLAGQPCADPRPREY
LILPEDCGQPGSGPAYLDVLCPTDTRCPSLAGSPRCGDASSCCSVRRLEREIHC
PGVLVPVKVVAECGCKCLPGLRGLRVVVAADSGEFLRFAILLGQEPFGFTAYQGD
FTTIEVPSTQRLVVTVFDPSPGEFMDAVRVLPFDPRGAGVYHEVKAMRKAPVILHTSQ
SNITPLGEDEAPLGBELVLPSPGAFRADQKPYSGPVEARVTFVDPRLDLSAASAPSD
LRVDSQDELAPLRTYGMFSVDLRAPGSARQLQGVPAVRVAASQIHMPGVRAALKW
SLNPETGLWEEESGFRREGSGPRVREERVELGNVEIRERRLFNDVPERRCFVK
ADRYANDKFTPSQVEGVVTVLNLEPGGFSANPRAWGRFDSAVTGPNGACLAFCFD
ADPDAAYTALVTATLGELEPAPSLPRPLPATVGTQPYLDRLGVRTDHDHPAFKR
NGFRINLAKPRPGDPAEANGFVYPMRSLRECQAPVTSYHFRFAREADKYEYNNVPF
REGTPASWTGDLIAWPNPOEFRAFLKVIQGPQEYMRVSHNAGGSHPRTRGQLYGL
RDARSVRDPERPGTSAACVEFKCSGMLFDQORVDRTLVTIMPOGSCRRVAVNGLRDY
LTRHPPVPAEDPAAFSMLAPLDPLGHNYGVTVTVDQSPRLAKETAIAGRCFDGSSDGF
SREMKADAGTAVTFOCREPPAGRPSLFQRLLESPTALGDIRREMSENAQAQARASGP
LRTRRGRVRQ"

ORIGIN
Query Match          90.6%; Score 15.4; DB 5; Length 3602;
Best Local Similarity 94.1%; Pred. No. 1.4e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCACGCTTGGCCGAG 17
||||| |||||||
Db 3256 GCCCACGCTTGGCCGAG 3272

RESULT 43
AR157945          4183 bp      DNA      linear      PAT 17-OCT-2001
LOCUS
DEFINITION      Sequence 2 from patent US 6251389.
ACCESSION      AR157945
VERSION      AR157945.1 GI:16219890
KEYWORDS      .
SOURCE      Unknown.
ORGANISM      Unknown.
REFERENCE      1 (bases 1 to 4183)
```

```

AUTHORS      Magna,H., Schaffer,P., Lawton,M., Yocum,S., Mitchell,P.,
              Hutchinson,N. and Murry L.E.
TITLE        Human nucleotide pyrophosphohydrolase-2
JOURNAL      Patent: US 6251389-A 2 26-JUN-2001;
FEATURES     Location/Qualifiers
              1..4183
              /organism="unknown"
              /mol_type="unassigned DNA"

ORIGIN
Query Match      90.6%; Score 15.4; DB 2; Length 4183;
Best Local Similarity 94.1%; Pred. No. 1.5e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGCCGAG 17
    |||||
Db 3254 GCCCAGCTTGCCCAAG 3270

RESULT 44
BD103585
LOCUS      BD103585              4183 bp      DNA      linear      PAT 27-AUG-2002
DEFINITION Human nucleotide pyrophosphohydrolase-2.
ACCESSION  BD103585
VERSION     BD103585.1 GI:22649159
KEYWORDS    JP 2001526886-A/1.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1 (bases 1 to 4183)
AUTHORS     Magna,H., Schaffer,P., Lawton,M., Yocum,S.A., Mitchell,P.G.,
            Hutchinson,N. and Murry,L.E.
TITLE       Human nucleotide pyrophosphohydrolase-2
JOURNAL     Patent: JP 2001526886-A 1 25-DEC-2001;
            INCYTE PHARMACEUTICALS INC
COMMENT     OS Homo sapiens (human)
            PN JP 2001526886-A/1
            PD 25-DEC-2001
            PF 02-DEC-1998 JP 2000525529
            PR 22-DEC-1997 US 08/996083
            PI HOLLY MAGNA,PAUL SCHAFER,MICHAEL LAWTON,SUE A YOCUM,PETER G
            MITCHELL,
            PI NANCY HUTCHINSON,LYNN E MURRY
            PC C12N9/16,A61K38/46,A61K39/395,A61K45/00,A61P19/02,A61P35/00,
            PC C07K16/40,C12N15/09,C12Q1/68,G01N33/15,G01N33/50,G01N33/53, PC
            A61K37/54,
            PC C12N15/00
            CC a or g or c or t, unknown, or other
            CC 1388013
            FH Key      Location/Qualifiers
            FT unsure   4180.
FEATURES     source
              1..4183
              /organism="Homo sapiens"
              /mol_type="genomic DNA"
              /db_xref="taxon:9606"

ORIGIN
Query Match      90.6%; Score 15.4; DB 2; Length 4183;
Best Local Similarity 94.1%; Pred. No. 1.5e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGCCGAG 17
    |||||
Db 3254 GCCCAGCTTGCCCAAG 3270

RESULT 45
CQ850573
LOCUS      CQ850573              4232 bp      DNA      linear      PAT 23-AUG-2004
DEFINITION Homo sapiens cDNA FLJ45835 fis, clone NTR29808057, weakly similar
            to Homo sapiens cartilage intermediate layer protein, nucleotide
            pyrophosphohydrolase (CILP).
ACCESSION  AK127735
VERSION     AK127735.1 GI:34534774
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
AUTHORS     Tanigami,A., Fujiwara,T., Shibahara,T., Goto,Y., Hirao,M.,
            Shimizu,F., Wakebe,H., Ono,T., Hishigaki,H., Watanabe,T., Ozaki,K.,
            Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S.,
            Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T.,
            Kimura,K., Yamashita,H., Matsuoka,K., Nakamura,Y., Sekine,M.,
            Kikuchi,H., Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K.,
            Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y.,
            Sugano,S., Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
            NEDO human cDNA sequencing project
            Unpublished
            2 (bases 1 to 4232)
            Isogai,T. and Yamamoto,J.
            Submitted (15-JUL-2003) Takao Isogai, Helix Research Institute,
            Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 293-0812, Japan
            (E-mail: flj-cdna@nifty.com, Tel.81-438-52-3975, Fax:81-438-52-3986)
            NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology (RAB); cDNA library
            construction: Helix Research Institute (HRI) (supported by Japan
            Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
            HRI, and Biotechnology Center, National Institute of Technology and
            Evaluation; clone selection for full insert sequencing: HRI and
            RAB; annotation: Reverse Proteomics Research Institute, HRI and
            RAB.
FEATURES     source
              1..4232
              Location/Qualifiers

```

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="NT2RP8008057"
/cell_line="NT2"
/cell_type="teratocarcinoma"
/clone_lib="NT2RP8"
/note="cloning vector: pME18SFL3
mRNA from NT2 neuronal precursor
retinoic acid (RA) induction."

ORIGIN

Query Match      90.6%; Score 15.4; DB 5; Length 4232;
Best Local Similarity 94.1%; Pred. No. 1.5e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17
Db 3323 GCCCAGCGTTGGCCCAAG 3339

RESULT 47
LOCUS AB209580 4641 bp mRNA linear PRI 31-MAR-2005
DEFINITION Homo sapiens mRNA for MHC class II antigen variant protein.
ACCESSION AB209580
VERSION AB209580.1 GI:62088739
KEYWORDS FLI CDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1
Totoki,Y., Toyoda,A., Takeda,T., Sakaki,Y., Tanaka,A., Yokoyama,S.,
Ohara,O., Nagase,T. and F.Kikuno,R.
None Title
PUBLISHED Only in Database (2005)
REFERENCE 2 (bases 1 to 4641)
Totoki,Y., Toyoda,A., Takeda,T., Sakaki,Y., Tanaka,A., Yokoyama,S.,
Ohara,O., Nagase,T. and F.Kikuno,R.
Direct Submission
Submitted (22-MAR-2005) Osamu Ohara, Kazusa DNA Research Institute,
Department of Human Gene Research; 2-6-7 Kazusa-Kamatari, Kisarazu,
Chiba, 292-0818, Japan (E-mail:cdnaif@kazusa.or.jp,
URL:http://protein.gsc.riken.go.jp/, Tel:81-438-52-3930,
Fax:81-438-52-3931)
This work was supported in part by the National Project on Protein
Structural and Functional Analysis, Ministry of Education, Culture
, Sports, Science and Technology of Japan. Totoki Y, Toyoda A,
Takeda T, Sakaki Y, Tanaka A, Yokoyama S. RIKEN Genomic Sciences
Center, 1-7-22 Suehiro-cho, Tsurumi, Yokohama 230-0045, Japan.
e-mail: aktanaka@postman.riken.go.jp
URL: http://protein.gsc.riken.go.jp/.
Location/Qualifiers
1. .4641
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="sj02456"
/tissue_type="spleen"
/note="this clone is also named as hek003001744
vector:pBluescriptII SK plus"
<1. .4641
/gene="MHC class II antigen variant"
<1. .499
/gene="MHC class II antigen variant"
/inference="non-experimental evidence, no additional
details recorded"
/note="Start codon is not identified."
/codon_start=2
/product="MHC class II antigen variant"
/protein_id="BAD92817.1"

gene
CDS

FEATURES
source
1. .4641
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/haplotype="DOB1*0602 DOA1*01021 DRB1*15011"
/cell_line="4AOH 100040 WAL FD"
<236. .>5425
/gene="HLA-DQB1"
/allele="HLA-DQB1*0602"
join(<236. .344,1776. .2045,4595. .4876,5386. .>5425)
/gene="HLA-DQB1"
/product="MHC class II antigen"
join(236. .344,1776. .2045,4595. .4876,5386. .>5425)
/gene="HLA-DQB1"
/allele="HLA-DQB1*0602"
/codon_start=1
/product="MHC class II antigen"
/protein_id="AAQ88044.1"
/db_xref="GI:37021371"
/translation="MSWKALRIPGDLRVATVTLMLAMLSLLAEGRSPEDFVFOFK
GMCYFTNGTERVRLVTRYINREYARFSDVGVRVTPQGRPDAEYVNSQKVELEG
TRAEFLDTCVRHNYEVAFRGILQRRGERRPSVRPTLGRDPFLCREGR
VWRFNDQEEETAGVYSTPLIRNGDWTTFQILVMLEMTFQRGDVYTCHVEHPHLSQSPITV
EWRAQSESAQSKMLSG"
<236. .344
/gene="HLA-DQB1"
/allele="HLA-DQB1*0602"
/number=1
1776. .2045
/gene="HLA-DQB1"
/allele="HLA-DQB1*0602"
/number=2

```

```

/db_xref="GI:62088740"
/translation="TTTAFPFVSMWKKALRIPGDLRVATVTLMLAMLSLLAEGRD
SPEDFVFOFKGMCYFTNGTERVRLVTRYINREYARFSDVGVRVTPQGRPDAEY
VNSQKVELEGTRAEFLDTCVRHNYEVAFRGILQRRGERRPSVRPTLGRDPFLCREGR
WGRGL"

ORIGIN

Query Match      90.6%; Score 15.4; DB 5; Length 4641;
Best Local Similarity 94.1%; Pred. No. 1.5e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17
Db 436 GCCCAGCGTTGGCCGAG 452

RESULT 48
LOCUS AY375870 5425 bp DNA linear PRI 31-AUG-2004
DEFINITION Homo sapiens 4AOH 100040 WAL FD MHC class II antigen (HLA-DQB1)
gene, HLA-DQB1*0602 allele, exons 1 through 4, and partial cds.
ACCESSION AY375870
VERSION AY375870.1 GI:37021370
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 5425)
Wood,J.M., Simons,M.J. and Ashdown,M.L.
Comparative haplotype analysis of polymorphism within the DQB1
gene. Continuous sequencing from promoter to exon 4
Unpublished
REFERENCE 2 (bases 1 to 5425)
Wood,J.M., Simons,M.J. and Ashdown,M.L.
Direct Submission
Submitted (25-AUG-2003) Genetype Pty Ltd, 60-66 Hanover Street,
Fitzroy, Victoria 3065, Australia
Location/Qualifiers
1. .5425
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/haplotype="DOB1*0602 DOA1*01021 DRB1*15011"
/cell_line="4AOH 100040 WAL FD"
<236. .>5425
/gene="HLA-DQB1"
/allele="HLA-DQB1*0602"
join(<236. .344,1776. .2045,4595. .4876,5386. .>5425)
/gene="HLA-DQB1"
/product="MHC class II antigen"
join(236. .344,1776. .2045,4595. .4876,5386. .>5425)
/gene="HLA-DQB1"
/allele="HLA-DQB1*0602"
/codon_start=1
/product="MHC class II antigen"
/protein_id="AAQ88044.1"
/db_xref="GI:37021371"
/translation="MSWKALRIPGDLRVATVTLMLAMLSLLAEGRSPEDFVFOFK
GMCYFTNGTERVRLVTRYINREYARFSDVGVRVTPQGRPDAEYVNSQKVELEG
TRAEFLDTCVRHNYEVAFRGILQRRVEPTVTPSPRTALNHNLLVCSVTDFPQGIK
VWRFNDQEEETAGVYSTPLIRNGDWTTFQILVMLEMTFQRGDVYTCHVEHPHLSQSPITV
EWRAQSESAQSKMLSG"
<236. .344
/gene="HLA-DQB1"
/allele="HLA-DQB1*0602"
/number=1
1776. .2045
/gene="HLA-DQB1"
/allele="HLA-DQB1*0602"
/number=2

```







```
/translation="MGNPSYNPPAGISPSDWLNLLOSAQRINPRPSPDFTDLKNYI
HWFHKTQKPTFTSGGPTSCPPGRFVPLVLAITNEVLSNEGAGCAGSAPEEOPPP
YDPAIPLIISSEGNRRHAWALRELQDIKEIENKARGSQWVITQLSLAIDQPTT
ADLEGCOYIASPVQDTAHTSLTAIAAENATPSRVLTPTKTLTQOASQPNAGDL
RSQYNLWLOAGKISLLVLQOPWSTIVOGPAESSVEFVRNMLQISLADNLDPGLRNP
LLTPLVMQLTSSVSKFCRGASGRGAKTAGLRTIGPPRMKQPALLVHTTFCPKMGP
RQAPAPRPDPGVRCLKEGHWARDCTKATGPPPGPCPCIKDPSHMKRDCPTLKSKN
"
```

mat\_peptide

628. .954

/gene="gag"

/product="gag-p15 polyprotein"

mat\_peptide

955. .1545

/gene="gag"

/product="gag-p24 polyprotein"

mat\_peptide

1597. .1803

/gene="gag"

/product="gag-12 polyprotein"

gene

2467. .4875

/gene="pol"

CDS

2467. .4875

/gene="pol"

/codon\_start=1

/product="pol protein"

/protein\_id="AAA42785.1"

/db\_xref="GI:210769"

```
/translation="VHDLRATNALTKPIPALSPGPDLTATPHPHIICLDLKDAPF
QIPVEDRFRVLSFTLSPGGLQPHREARVLPQGFINSALFERALQELRQVSAA
FSQSLVSYMDILYASTERQSCYOALARNLDLGFVASEKTSITPSPVFLQG
MVRHQIYTSQLTQISPSLHQLQAVLDGQVWSRGTTPTPRLQLLYSLKRRH
DPRAIQLSPQLOAGIALRQALSHNARSRYNEQELVYVHLTRAGSLVLFQKGAQ
FPYAFOTPLDQNASPWGLLLGCOVLOQALSSYAKPIKYHNLPKTSLDNWIO
SSEPRVQLLQLOPISOSGLOPPGPKTLITRAEVFLTPQFSDPTPAALCLFSDG
ATGRGACLWKLHLLDFOAVPAPSAOKGELAGLAGLAAPPEPNIWDSKLYLSL
LRTVLGAWLQDPVPVSVALYKSLLRHPALVGVHRSASPSIASLNNYVDQLLP
LETPEQMKLTHCNSRLSRWPNRISADPSRPAIETCETCKLNPTGGGKMTIQRG
WAPNHIOADITHYKQFTYALHVFVDVTSYGHASAKRGLTTQTTIEGLEAIVHL
GRPKNLDOQANTYKTFVRFQCGVSLSHVYPNPTSSGLDERTNGLLKLILSKY
HLDEPLPMTQALSRALWTHQINLLPIKTRWELHSPPLAVISEGGETPKGSDKLF
LHLLPGQNNRRWLGLPALVEASGALLATDPVWPVWRLKAKFKCLKNDGPDAHNR
SSDG"
```

gene

4821. .6368

/gene="env"

CDS

4821. .6368

/gene="env"

/codon\_start=1

/product="envelope polyprotein"

/protein\_id="AAA42786.1"

/db\_xref="GI:210770"

```
/translation="MPKERRRRPQPIIRWVSLTLTLLALCRPIQTWRCSLSLGNQO
WMTAYNOEKFSGIDQILEAHNOSPFCAPRYTLDVNGVYPKIYWPPOGRRRFGA
RAMVYDCEPCPYGADRFDCPHDNASQADQSFYVNHQILFLHLKQCHGIFLTW
EINGYDLITSLHKIDPPQDPQLNSDWVSVRSWALLNQATAPPCALCWEPE
SPWAPETLVNKTISSGGGLADPDAQIFWNSSSNTTQGHHPSPORLFFNVSQGN
ALLPLSLTSLNASAPPTRRSPVAALTLGLSVGLTGINAVSALSHQRLTS
LIHVLEQDITAINQTHYNLLNVAQVNRRLDMLYIRLGFOSLCPCTINEPCC
FLRLQNDQIILRLGLOPLOSQVRSQVTDQWPNNDLGTAWVRETIHSLVSLFLALFLL
FLAPCLIKLTSRLKLLRQAPFPEISLTPKPSDQVALLPSAPEIYSHLSVPKPDY
INLRPCP"
```

sig\_peptide

4821. .4919

/gene="env"

mat\_peptide

4920. .5723

/gene="env"

mat\_peptide

5724. .6365

/product="gp51 surface glycoprotein (env)"

/gene="env"

/product="gp30 transmembrane protein (env)"

8185. .8714

/standard\_name="3' LTR"

/db\_xref="P01111"

ORIGIN

63 bp upstream of PvuII site.

Query Match 90.6%; Score 15.4; DB 10; Length 8714;

Best Local Similarity 94.1%; Pred. No. 1.6e+03;

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY 1 GCCCAGCGCTTCCCGAG 17
|||||
Db 4288 GCCCAGCGCTTCCCGAG 4272
```

RESULT 53

AE004474/c

LOCUS

DEFINITION

Pseudomonas aeruginosa PAO1, section 35 of 529 of the complete genome.

ACCESSION

AE004474

VERSION

AE004474.1

KEYWORDS

SOURCE

ORGANISM

Pseudomonas aeruginosa PAO1

Pseudomonas aeruginosa PAO1

Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

Pseudomonadaceae; Pseudomonas.

REFERENCE

AUTHORS

Stover, C.K., Pham, X.-Q.T., Erwin, A.L., Mizoguchi, S.D., Warren, P., Hickey, M.J., Brinkman, F.S.L., Hufnagle, W.O., Kowalik, D.J., Lagrou, M., Garber, R.L., Goltry, L., Tolentino, E., Westbrock-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Larbig, K., Lim, R.M., Smith, K.A., Spencer, D.H., Wong, G.K.-S., Wu, Z. and Paulsen, I.T.

Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen

NATURE 406 (6799), 959-964 (2000)

JOURNAL

PUBMED

REFERENCE

AUTHORS

Stover, C.K., Pham, X.-Q.T., Erwin, A.L., Mizoguchi, S.D., Warren, P., Hickey, M.J., Brinkman, F.S.L., Hufnagle, W.O., Kowalik, D.J., Lagrou, M., Garber, R.L., Goltry, L., Tolentino, E., Westbrock-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Larbig, K., Lim, R.M., Smith, K.A., Spencer, D.H., Wong, G.K.-S., Wu, Z., Paulsen, I.T., Reizer, J., Saier, M.H., Hancock, R.E.W., Lory, S. and Olson, M.V.

Direct Submission

Submitted (16-MAY-2000) Department of Medicine and Genetics, University of Washington Genome Center, University Of Washington, Box 352145, Seattle, WA 98195, USA

3 (bases 1 to 10694)

Pseudomonas aeruginosa Community Annotation Project (PseudocAP)

Direct Submission

Submitted (04-FEB-2003) Department of Molecular Biology and Biochemistry, Simon Fraser University, 8888 University Dr., Burnaby, British Columbia V5A 1S6, Canada

TITLE

JOURNAL

REFERENCE

AUTHORS

CONSTRM

TITLE

JOURNAL

COMMENT

Pseudomonas aeruginosa PAO1 genome annotation, updated, reviewed, Pseudomonas aeruginosa PAO1 genome annotation, from PseudocAP (see <http://www.pseudomonas.com> for latest updates and links to alternate annotations). PseudocAP is coordinated by Fiona S.L. Brinkman (Simon Fraser University, Canada) and Robert E.W. Hancock (University of British Columbia, Canada). We welcome submission through [www.pseudomonas.com](http://www.pseudomonas.com) of any proposed changes.

Protein name confidence' is used to rate our confidence of the accuracy of the protein name.

Class 1: Function experimentally demonstrated in P. aeruginosa.

Class 2: Function of highly similar gene experimentally demonstrated in another organism (and gene context consistent in terms of pathways its involved in, if known).

Class 3: Function proposed based on presence of conserved amino acid motif, structural feature or limited sequence similarity to an experimentally studied gene.

Class 4: Homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences.

Location/Qualifiers

1. .10694

/organism="Pseudomonas aeruginosa PAO1"

/mol\_type="genomic DNA"

```
/strain="PA01"
/db_xref="taxon:208964"
166..1152
/locus_tag="PA0360"
166..1152
/locus_tag="PA0360"
/notes="Protein name confidence: Class 4 (homologs of
previously reported genes of unknown function, or no
similarity to any previously reported sequences)"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG03749.1"
/db_xref="GI:9946211"
/transl_table=11
/translation="MFSQWVFLSVLALVPSFAWAEDAVSDLAREQSKAEQDKRGDE
LFAEFLRQAAETGEPEQALDFGAFYQQNTYASAREWKGAAAGMVRQAQLAML
YRGDGGPODKTAARWFERKAAEOGDAAGNMGVLVWYRGEGVDQDRVKAGTWFERAA
AGSEDAETNLGHFYLDDSQVYSATSDDEANLLDRYAGSREKAFOWECKANTQGDAR
AQKVGAYWNGSGAGWNKQARLWLEKAAQQQADAIWLESNANAATWTRLENWTH
AALDGEUGVADCSGAADRRHSGRGMGEPSPDPEAVEASAEADAQQAVEAASQ"
/locus_tag="PA0361"
/complement(1257..2990)
/locus_tag="PA0361"
/complement(1257..2990)
/locus_tag="PA0361"
/notes="Protein name confidence: Class 3 (function proposed
based on presence of conserved amino acid motif,
structural feature or limited sequence similarity to an
experimentally studied gene); Subcellular localization:
Periplasmic; Subcellular localization confidence: Class 2"
/codon_start=1
/transl_table=11
/product="probable gamma-glutamyltranspeptidase precursor"
/protein_id="AAG03750.1"
/db_xref="GI:9946212"
/transl_table=11
/translation="MSFLNLPVRTLVAACGLSLSLAAHADSPRGDAHAPQAAVATP
HPAATVAGLETLANGNADFADAAIAALAAVAEPYSGLGSGGPFLLRQAGOPTYFF
LDARERAPKADYMYRNKGVDPRLSDVGLPAAAIKGLPALVELSRVGRKPLADN
LVPAIRLAVGVSDYRDRRAARLEAMRKDPETARIFLDKGGIPDENLLRQPLA
RTLRLGRYGRIGYEGETAOKLTLGVYRAGGIMSAADLRDYRIERRPLEVKLANGR
ELLSAPPPSAGGVAAQLEOLPWQKAEFPVQARHYVLEVLRYARDRGLGLDPPF
VANPLSLAPDYLRKLAAGIDPRRATPSSALPEAPAWREGDHTTHEAVIDAQGNVA
ATLSVNLPGCAFTVPTGVNLMDDFADTCTOGANSYGLAGSQANAAAGKRPGLS
MSPFLSPDTPFAAFGTGGSRIFSMVLISMLQYFDGRPVQGWAAAPRHHQYKPDV
EYEPRAFSDAEAEALRRRGYQLKRLERSYGNQQVLYWKRSGRVDAAASDRFGYISL
SLEAQAPRPH"
/complement(3118..3369)
/gene="fdx1"
/locus_tag="PA0362"
/complement(3118..3369)
/gene="fdx1"
/locus_tag="PA0362"
/notes="Protein name confidence: Class 2 (high similarity
to functionally studied protein)"
/codon_start=1
/transl_table=11
/product="ferredoxin [4Fe-4S]"
/protein_id="AAG03751.1"
/db_xref="GI:9946213"
/translation="MSLKITDDCTNCDCVEPCPNAGISQGEIYVIDPNLCTECVGH
YDEPCCQVCPVDCIPLDANVESKDQLMEKYRITGKA"
/complement(3490..3969)
/gene="coad"
/locus_tag="PA0363"
/notes="Protein name confidence: Class 2 (high similarity
to functionally studied protein); Alt protein name: PPA1"
/codon_start=1
/transl_table=11
/product="phosphopantetheine adenylyltransferase"
/protein_id="AAG03752.1"
```

```
/db_xref="GI:9946214"
/translation="MNRVLYPGTFDPTTKGHGDLIERASRLFDHVIIVAAASPKKNPL
FSLEQRVALAQEVTKHLPNVEVVGFTLLAHFVKQKQANVFLRLGRVAFDFEYFQLA
NMNRQAPDVESMFLTPSEKYSFISSTLVREIAALGGDISKFVHPAVADALAEFRK"
/complement(4121..5716)
/locus_tag="PA0364"
/complement(4121..5716)
/locus_tag="PA0364"
/notes="Protein name confidence: Class 3 (function proposed
based on presence of conserved amino acid motif,
structural feature or limited sequence similarity to an
experimentally studied gene)"
/codon_start=1
/transl_table=11
/product="probable oxidoreductase"
/protein_id="AAG03753.1"
/db_xref="GI:9946215"
/transl_table=11
/translation="MPVPDLFAEGLRAGWKTYNGSRLLDDLEADVVVVGAGGGT
TAEILSAGLKVLLEVEGPKTSDFKQWQADAYPELYQEGIGRMSKDGATITLQGRA
VGGTTLINSSFTPEPTLQHWAOYVGVKGHSAEDMAPFEKQEORLHVAPWALPEN
ANNVDRLGCEKLGHWKVIIPRNVLGWNLCYGMGCPVNAKQSLMTITLIPATLDKGG
ELLYLARANRLLDGDKVTGELGMDERCVPNRRIRVRARHYVLSGGGINTPAIL
LRKAPDPQSQRVGRKTPHTWFSAGLFDVINPFYGAQSIYSDHFQWDDGVTRMS
KLEVPPLQSLASVLLGGFGSDNALRMEOLPHTNMMLALLRDCDFHDPDSARGSVLGRG
DTPPLDYRTDVTMDGIRAFHTMAEIOFAAGAREVPVHSDARSANLQEAEGIID
GLREIYTRTSGAHVMGGCAGEDPAVAVADSLGRHHLNLSIHGDSLPTSIGAN
PQUSVGLTAQLASALAERLARA"
/complement(5772..6320)
/locus_tag="PA0365"
/complement(5772..6320)
/locus_tag="PA0365"
/notes="Protein name confidence: Class 4 (homologs of
previously reported genes of unknown function, or no
similarity to any previously reported sequences)"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG03754.1"
/db_xref="GI:9946216"
/translation="MSDTTLESAGLSRRSLMKVGLIGAFATAGVTASLTGCSAEKP
ASGLKVRSDPLPRLALLPVLILGAVSAEQMPKAVEGAIOSLDHNLARLSPMPKLT
QQLFDVLVALPTRGPLTGIWGSWENASGDDVRAFLSRWENSFIGLLRMGHSSLMQLAM
MAYARPEANAHCGYPGPKIA"
/complement(6375..7805)
/locus_tag="PA0366"
/complement(6375..7805)
/locus_tag="PA0366"
/notes="Protein name confidence: Class 3 (function proposed
based on presence of conserved amino acid motif,
structural feature or limited sequence similarity to an
experimentally studied gene)"
/codon_start=1
/transl_table=11
/product="probable aldehyde dehydrogenase"
/protein_id="AAG03755.1"
/db_xref="GI:9946217"
/translation="MVADIAVLYQSQEQEINQLEALFAQRQAAYLAQPMPDATQRVQWL
KALRDLFLKQEQALIEADRDFSARSADETLLAIMPSLGHIGHAAKRVKWKMPARR
AVGLQFQPAQAQVYQPLVGVVIVPWNYPLFLSIGPLTGALAAGRNVMYIKSESTPA
TGRLLKLLARIIPEDQVAVVGVAVFASPKLPFDHLLFTGATSVGHVWRAAEEN
LTPVTLELGGKSPAIVSDSPVMKDAERIAFGKSLNAGOTCVADPYLVPSRRVREFV
SQYKEVVQGFPPRLSDNPDYTAIINERQLGRGLYLDAREKGATLVFLFAEQQRRL"
Query Match 90.6%; Score 15.4; DB 15; Length 10694;
Best Local Similarity 94.1%; Pred. No. 1.6e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 GCCCACGCTTCGCCGAG 17
|||||
Db 7832 GCCCACGCTTCGCCGAG 7816
RESULT 54
```



Query Match 90.6%; Score 15.4; DB 6; Length 59316;  
Best Local Similarity 94.1%; Pred. No. 2e+03;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCACGCTGGCCGAG 17  
DB 38622 GCCACGCTGGCCGAG 38638

RESULT 56  
AC165982  
LOCUS AC165982 Bos taurus clone CH240-173K8, 66661 bp DNA linear HTG 20-JUN-2005  
DEFINITION Bos taurus clone CH240-173K8, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 20  
unsorted pieces.  
ACCESSION AC165982.1 GI:71037423  
VERSION HTG; HTGS PHASE1  
KEYWORDS Bos taurus (cattle)  
SOURCE Bos taurus  
ORIGINISM Bos taurus

REFERENCE  
AUTHORS Muzny,D., Adams,C., Agbai II,O., Allen,C., Alsbrooks,S., Archer,P.,  
Arredondo,H., Bandaranaike,D., Bangura,L., Beltran,B., Beltran,R.,  
Beraducci,A., Biswal,K., Blyth,P., Bonham,H., Buhay,C., Burch,P.,  
Cadoree,I., Canada,A., Cardenas,V., Carter,K., Cavazos,I.,  
Chacko,J., Chahrour,M., Chavez,D., Chen,A., Chen,G., Chen,R.,  
Cheng,M.-T., Chu,J., Clerc,K., Cockrell,R., Coyle,M., Cree,A.,  
Curry,S., Dai,W., Davila,M.L., Davis,C., Davy-Carroll,L., De  
Anda,C., Delgado,O., Denson,S., Dugan-Rocha,S., Dunn,A., Durbin,K.,  
Donlin,J., McCauley,S., Dugan-Rocha,S., Dunn,A., Durbin,K.,  
DiLuia,D., Egan,A., Escotto,M., Espinosa,V., Eugene,C., Fa,M.,  
Fernandez,S., Fernandez,P., Flagg,N., Forbes,L., Foster,P.,  
Fowler,G., Fu,Q., Fuh,E., Garcia,A., Garcia,R., Garner,T.,  
Gaskin,C., Gensch,S., Ghose,S., Gill,R., Gonzalez,D.,  
Gonzalez-Garay,W., Guevara,W., Holder,M., Haaland,W., Haerberlen,K.,  
Hall,B., Hamid,H., Hamilton,K., Harbes,B., Harris,R., Havlak,P.,  
Hawes,A., Hawkins,E., Hayes,S., Hemphill,L., Hernandez,J.,  
Hines,S., Hitchens,M., Hodgson,A., Hognes,M., Hollins,B.,  
Howell,L.T., Hulyk,S., Hume,J., Imo,K., Jackson,A., Jackson,L.,  
Jacob,L., Jiang,H., Johnson,B., Johnson,R., Kalafus,K., Kelly,S.,  
Keys,T., Khan,Z., King,L., Kovar,C., Kowis,A., Kowis,C., Lara,F.,  
Leal,S., Lee,K., Lee,S., Legall,F.I., Lemon,S., Lewis,L., Li,B.,  
Li,Y., Li,Z., Linnell,M., Liu,W., Liu,Y.-S., Liu,Y., Liyanage,D.,  
London,P., Lopez,J., Lorensuhewa,L., Lozado,R., Luk,T., Madu,R.,  
Maheshwari,M., Mahoney,C., Malloy,K., Mansouri,D., Martinez,E.,  
McClelland,H., McPherson,J., Mercadao,C., Metzker,M.,  
Mlivosavljevic,A., Minja,E., Morgan,M., Morris,S., Munidasa,M.,  
Murray,D., Nazarith,L., Ngo,D., Nguyen,N., Norwig-Eastaugh,E.,  
Nott,A., Nwaelelemeh,O., Obregon,M., Ochi-Okorie,C., Odeh,E.,  
Okwuonu,G., Okwuonu,K., Parker,D., Pasternak,S., Patel,B.,  
Patel,V., Paul,H., Perez,A., Perez,L., Petrosino,J., Pham,T.,  
Primus,E., Pu,L.-L., Puazo,M., Qin,X., Quinn,A., Quiroz,J.,  
Rabata,D., Rachin,E., Reigh,R., Ren,Y., Reuter,M., Richards,S.,  
Rives,C., Rodriguez,F., Rojas,A., Ruiz,S.J., Sana,M., Sanders,W.,  
Santibanez,J., Santos,R., Savary,G., Scherer,S., Shen,H., Shen,Y.,  
Sisson,A., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R.,  
Statok,A., Taylor,E., Taylor,T., Thomas,N., Thorn,R., Thornton,R.,  
Trejos,Z., Ugarni,K., Vargo,C., Verduzco,D., Villasana,D., Virk,D.,  
Volkov,A., Waldron,L., Walker,B., Wang,Q., Wang,S., Warren,J.,  
Wei,X., Wheeler,D., Williams,G., Williams,R., Worley,K., Wright,R.,  
Wu,J., Yakub,S., Yan,K., Yuan,Y., Yu,F., Zhang,J., Zhang,L.,  
Zhang,Z., Zhou,J., Weinstein,G. and Gibbs,R.

TITLE Direct Submission  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 56661)  
AUTHORS Worley,K.C.  
TITLE Direct Submission  
JOURNAL Submitted (20-JUL-2005) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
----- Genomes Center

Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help.bcm.tmc.edu  
----- Project Information  
Center project name: FKEV  
Center clone name: CH240-173K8  
----- Summary Statistics  
Sequencing vector: Plasmid;  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 68288 bases at least Q40  
Consensus quality: 73110 bases at least Q20  
Estimated insert size: 94571; sum-of-contigs estimation  
Quality coverage: 1x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 20 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\*  
\* 1 3787: contig of 3787 bp in length  
\* 3788: gap of unknown length  
\* 3887: contig of 2287 bp in length  
\* 6174: gap of unknown length  
\* 6274: gap of unknown length  
\* 6175  
\* 6275: contig of 3914 bp in length  
\* 10189  
\* 10288: gap of unknown length  
\* 12369  
\* 12368: contig of 2080 bp in length  
\* 12369  
\* 12468: gap of unknown length  
\* 12469  
\* 12480: contig of 4812 bp in length  
\* 17281  
\* 17380: gap of unknown length  
\* 20213: contig of 2833 bp in length  
\* 20313: gap of unknown length  
\* 25626: contig of 5313 bp in length  
\* 20314  
\* 25626: gap of unknown length  
\* 25627  
\* 25726: gap of unknown length  
\* 25727  
\* 27874: contig of 2148 bp in length  
\* 27875  
\* 27974: gap of unknown length  
\* 30254: contig of 2280 bp in length  
\* 30255  
\* 30354: gap of unknown length  
\* 33291: contig of 2937 bp in length  
\* 33291  
\* 33291: gap of unknown length  
\* 33392  
\* 33927: contig of 5436 bp in length  
\* 38828  
\* 38928  
\* 41649  
\* 41748: gap of unknown length  
\* 41749  
\* 44783: contig of 3035 bp in length  
\* 44883: gap of unknown length  
\* 46889: contig of 2006 bp in length  
\* 46890  
\* 46890: gap of unknown length  
\* 46909  
\* 49308: contig of 2319 bp in length  
\* 49409  
\* 51454: gap of unknown length  
\* 51454  
\* 51555: gap of unknown length  
\* 51555  
\* 55173: contig of 3519 bp in length  
\* 55074  
\* 55174: contig of 3141 bp in length  
\* 58315  
\* 58414: gap of unknown length  
\* 58415  
\* 62032  
\* 62131: gap of unknown length  
\* 62132  
\* 66661: contig of 4530 bp in length.  
\*  
\* Location/Qualifiers  
\* 1..66661  
\* /organism="Bos taurus"  
\* /mol\_type="genomic DNA"  
\* /db\_xref="taxon:9913"  
\* /clone="CH240-173K8"  
\* 3788..3887  
\*  
\* FEATURES  
\* source  
\* gap

/estimated\_length=unknown  
 6175. .6274  
 /estimated\_length=unknown  
 10189. .10288  
 /estimated\_length=unknown  
 12369. .12468  
 /estimated\_length=unknown  
 17281. .17380  
 /estimated\_length=unknown  
 20214. .20313  
 /estimated\_length=unknown  
 25627. .25726  
 /estimated\_length=unknown  
 27875. .27974  
 /estimated\_length=unknown  
 30255. .30354  
 /estimated\_length=unknown  
 33292. .33391  
 /estimated\_length=unknown  
 38828. .38927  
 /estimated\_length=unknown  
 41649. .41748  
 /estimated\_length=unknown  
 44784. .44883  
 /estimated\_length=unknown  
 46890. .46989  
 /estimated\_length=unknown  
 49309. .49408  
 /estimated\_length=unknown  
 51455. .51554  
 /estimated\_length=unknown  
 55074. .55173  
 /estimated\_length=unknown  
 58315. .58414  
 /estimated\_length=unknown  
 62032. .62131  
 /estimated\_length=unknown

## ORIGIN

Query Match 90.6%; Score 15.4; DB 12; Length 66661;

Best Local Similarity 94.1%; Pred. No. 2.1e+03;

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGGCCGAG 17

Db 29600 GCCACGCTTGGCCGAG 29616

RESULT 57  
 AY663411  
 LOCUS  
 DEFINITION  
 Homo sapiens voucher Coriell Cell Repository DNA sample NA14663 MHC  
 class II antigen (HLA-DQB1) gene, HLA-DQB1\*0602 allele, and  
 class II antigen (HLA-DQAI) gene, HLA-DQAI\*010201 allele, and  
 MHC class II antigen (HLA-DRB1) gene, complete cds.

ACCESSION AY663411

VERSION AY663411.1 GI:52840236

KEYWORDS

SOURCE

ORGANISM

Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

REFERENCE 1 (bases 1 to 99966)

AUTHORS

Raymond, C.K., Kas, A., Paddock, M., Qiu, R., Zhou, Y., Subramanian, S.,

Chang, J., Palmieri, A., Haugen, E., Kaul, R. and Olson, M.V.

TITLE

Ancient haplotypes of the HLA Class II region

JOURNAL

Genome Res. 15 (9), 1250-1257 (2005)

PIUMED

16140993

REFERENCE 2 (bases 1 to 99966)

AUTHORS

Raymond, C.K., Kas, A., Paddock, M., Qiu, R., Zhou, Y., Subramanian, S.,

Chang, J., Palmieri, A., Haugen, E., Kaul, R. and Olson, M.V.

TITLE

Direct Submission

JOURNAL

Submitted (18-JUN-2004) Genome Center, Department of Medicine,  
University of Washington, Box 352145, Seattle, WA 98195, USA

FEATURES

source

Location/Qualifiers

1. .99966

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/specimen\_voucher="Coriell Cell Repository DNA sample

NA14663"

/db\_xref="taxon:9606"

<3025. .>9639

/gene="HLA-DQB1"

/alleles="DQB1\*0602"

join(<3063. .3171.4603. .4872.7422. .7703.8213. .8323,

9421. .>9434)

/gene="HLA-DQB1"

/allele="DQB1\*0602"

/product="MHC class II antigen"

join(3063. .3171.4603. .4872.7422. .7703.8213. .8323,

9421. .9434)

/gene="HLA-DQB1"

/allele="DQB1\*0602"

/codon\_start=1

/product="MHC class II antigen"

/protein\_id="AAU88024.1"

/db\_xref="GI:52840239"

/translation="MSWKALRIKGLRVATVTLMLSSLAEGRDSDPEDVFOFK

CMCYFTNGTERVRLVRYIYNREYAREDSVGVYRAVTPQGRDAEYWNQKEVLEG

TRALDTVCRHNYEVAFCILQRRVEPTVITSPSTEALNHNHLLVCSTVDYDYGQIK

VWFRNDQETAGVSTPLIRNGDWFQILVNLMTPOQGDVYTCHEVHPSLQSPITY

EWRAQSESAQSKMLSGVGGFVLGLIFLGLGLIIRQSQKGLLH"

complement (26077. .32251)

/gene="HLA-DQAI"

/allele="DQAI\*010201"

complement (join(<26903. .27057.27414. .27695.28109. .28357,

32128. .>32209))

/gene="HLA-DQAI"

/alleles="DQAI\*010201"

/product="MHC class II antigen"

complement (join(26903. .27057.27414. .27695.28109. .28357,

32128. .32209))

/gene="HLA-DQAI"

/alleles="DQAI\*010201"

/codon\_start=1

/product="MHC class II antigen"

/protein\_id="AAU88022.1"

/db\_xref="GI:52840237"

/translation="MILNKALLGALATTWNSPCGGEDIVADHVASCQVNLVYQFVGP

SGQYTHEFDGEQYVDLERKETAWRWEFSKFGGDFQGALRNMAVAKHNINIMIKR

YNSTAATNEVPEVTVFSKSPVTLGQPNLILCLVDNIFPPVWNITWLSNGQSVTEGVSE

TSFLSKSDHSFFKISYLTFLPSADEIYDKVEHMGLDQPLLKHWEPEIPAPNSELTET

VVCAIGLSVGLMGIVGVTFIIQGLRSVGSRRHQGPL"

<79861. .>90906

/gene="HLA-DRB1"

join(<79937. .80036.85301. .85570.87837. .88118.88819. .88929,

89405. .89428.90571. .>90584)

/gene="HLA-DRB1"

/product="MHC class II antigen"

join(79937. .80036.85301. .85570.87837. .88118.88819. .88929,

89405. .89428.90571. .90584)

/gene="HLA-DRB1"

/codon\_start=1

/product="MHC class II antigen"

/protein\_id="AAU88023.1"

/db\_xref="GI:52840238"

/translation="MVCUKLPGSCMTATVTLMLVSSFLASGSTRFRRLWQPKREC

HFNGTERVRLDRHNYEVSVPDSVDFEFAVTELGREDAEYWNQKDLQEARA

AVDTYCRHNYGVESFVQRRVQPKVTVYPSKTPQLQHNHLLVCVSGFYPGSIEVRW

FLNGEEKAGMVTGLIQNGDWTFTLLVLETPRSGEYVTCOVHPSVTSPLTVEWR

ARSESQAQSKMLSGVGGFVLGLIFLGLGLFIIFRNQKHSGHQPTGFLS"

## ORIGIN

Query Match

Best Local Similarity 90.6%; Score 15.4; DB 5; Length 99966;

94.1%; Pred. No. 2.2e+03;

```

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCACCCTTGCCCGAG 17
||||| |||||||
Db 4898 GCCCACCCTTGCCCGAG 4914

RESULT 59
AY663415 100761 bp DNA linear PRI 08-SEP-2005
LOCUS HSDJ93P18/103216 bp DNA linear PRI 18-MAY-2005
DEFINITION Human DNA sequence from clone RPI-93P18 on chromosome 1p35.2-36.23
Contains the 5' end of the ALDH4A1 gene for aldehyde dehydrogenase
4 family, member A1, a novel gene and a CpG island, complete
sequence.
ACCESSION AL080251
VERSION AL080251.23 GI:15986414
KEYWORDS HTG; ALDH4A1; CpG island.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE Moore.M.
AUTHORS Direct Submission
TITLE Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
JOURNAL Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
COMMENT On Oct 8, 2001 this sequence version replaced gi:15963342.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep/
This sequence
was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chrl
RPI-93P18 is from the library RPI-1 constructed by the group of
Pietter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2
----- Genome Center
Center: Wellcome Trust Sanger Institute

/protein_id="AAU88034.1"
/db_xref="GI:52840253"
/translation="MILNKALLIGALALTTVMSPCGGEDIADHVASCGVNLQFYQGP
SGQTFHEFDGDEQFYVDLEKETAWRWPEFSKFGGFDPOGALRNMVAKHNLNMIKR
YNSTAATNEVEPTVFSKSPVTLGQPNLTICLVNDI PPVNVNITLWNGHAVTEGVSE
TSLPSKSDHSFKIISYLTFLPSADELYDCKVEHWGLDQPLLKHWEPI PAMPSELTE
VVCALGSLVGLGIVGVTFIIQGLRSVGASRHQGPL"
<80351..>93754
/gene="HLA-DRB1"
/join(<80427..80526,88518..88787,91040..91321,92020..92130,
92606..92629,93418..>93431)
/gene="HLA-DRB1"
/product="MHC class II antigen"
/join(80427..80526,88518..88787,91040..91321,92020..92130,
92606..92629,93418..93431)
/gene="HLA-DRB1"
/codon_start=1
/product="MHC class II antigen"
/protein_id="AAU88035.1"
/db_xref="GI:52840254"
/translation="MVCRLRPGGSCMAVLTVTLMLVSSPLALAGDTTRPRFLEYSTSEC
HFFNGTRVRFLDRYFHNQENVRVFDSDVGEFRAVTELGFRDAEYMWNSQKIDILDERA
AVDTYCEHNYGVVESFTVQRVHPKTVPSKTOPLOHNLVLCVSGFYGSGIIEVRW
FRNGQEKTVGVSTGLIHNGDWTFOETLVMLETVPKRGVVTTCVEHPSVTSPLTVEWR
ARSEASQKMLSGVGGFVLGLLFLGAGLFIYFRNQKHSGLQPRGFLS"

ORIGIN
Query Match 90.6%; Score 15.4; DB 5; Length 100761;
Best Local Similarity 94.1%; Pred. No. 2.2e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCACCCTTGCCCGAG 17
||||| |||||||
Db 9393 GCCCACCCTTGCCCGAG 9409

RESULT 59
HSDJ93P18/c
LOCUS HSDJ93P18/103216 bp DNA linear PRI 18-MAY-2005
DEFINITION Human DNA sequence from clone RPI-93P18 on chromosome 1p35.2-36.23
Contains the 5' end of the ALDH4A1 gene for aldehyde dehydrogenase
4 family, member A1, a novel gene and a CpG island, complete
sequence.
ACCESSION AL080251
VERSION AL080251.23 GI:15986414
KEYWORDS HTG; ALDH4A1; CpG island.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE Moore.M.
AUTHORS Direct Submission
TITLE Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
JOURNAL Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
COMMENT On Oct 8, 2001 this sequence version replaced gi:15963342.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep/
This sequence
was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chrl
RPI-93P18 is from the library RPI-1 constructed by the group of
Pietter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2
----- Genome Center
Center: Wellcome Trust Sanger Institute

/gene
mRNA
CDS

AY663415 100761 bp DNA linear PRI 08-SEP-2005
Homo sapiens voucher Coriell Cell Repository DNA sample NA03715 MHC
Class II antigen (HLA-DQB1) gene, MHC class II antigen (HLA-DQA1)
gene, HLA-DQA1-DQA1*0103 allele, and MHC class II antigen
(HLA-DRB1) gene, complete cds.
AY663415
AY663415.1 GI:52840252
Homo sapiens (human)
Homo sapiens
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 100761)
Raymond,C.K., Kas,A., Paddock,M., Qiu,R., Zhou,Y., Subramanian,S.,
Chang,J., Palmieri,A., Haugen,E., Kaul,R. and Olson,M.V.
Ancient haplotypes of the HLA Class II region
Genome Res. 15 (9), 1250-1257 (2005)
16140993
2 (bases 1 to 100761)
Raymond,C.K., Kas,A., Paddock,M., Qiu,R., Zhou,Y., Subramanian,S.,
Chang,J., Palmieri,A., Haugen,E., Kaul,R. and Olson,M.V.
Direct Submission
Submitted (18-JUN-2004) Genome Center, Department of Medicine,
University of Washington, Box 352145, Seattle, WA 98195, USA
Location/Qualifiers
1..100761
/organism="Homo sapiens"
/mol_type="genomic DNA"
/specimen_voucher="Coriell Cell Repository DNA sample
NA03715"
/db_xref="taxon:9606"
<7520..>14135
/gene="HLA-DQB1"
/join(<7558..7666,9098..9367,11918..12199,12709..12819,
13917..>13930)
/gene="HLA-DQB1"
/product="MHC class II antigen"
/join(7558..7666,9098..9367,11918..12199,12709..12819,
13917..13930)
/gene="HLA-DQB1"
/codon_start=1
/product="MHC class II antigen"
/protein_id="AAU88036.1"
/db_xref="GI:52840255"
/translation="MSWKALKRIFGDLRVATVTLMLSLAEGRDSPEDFVYQFK
GMCFYTGTRVRLVTHYINREYARFDSVGYRAVTPQGRPDAYWNSQKVLGK
TRALDVTCHRYNEVAFRGLIQREYVPTVTSRTEALNHNLLVCSVTDFYQGIK
VWRFNDOEETAGVSTPLIRNGDWTFOETLVMLETVPKRGVVTTCVEHPSLQSPITV
EWRAQSSAQKMLSGVGGFVLGLIIFLGGLIIRQSQKGLLH"
complement(30616..36797)
/gene="HLA-DQA1"
/allele="DQA1*0103"
complement(join(<31442..31596,31953..32234,32648..32896,
36674..>36755))
/gene="HLA-DQA1"
/allele="DQA1*0103"
/product="MHC class II antigen"
complement(join(31442..31596,31953..32234,32648..32896,
36674..36755))
/gene="HLA-DQA1"
/allele="DQA1*0103"
/codon_start=1
/product="MHC class II antigen"

```





```
complement(AL954340.4:2548..2701),
complement(BX537160.6:16782..17140))
/genes="ALDH4A1"
/locus tag="RPL1-128M10.1-002"
/product="aldehyde dehydrogenase 4 family, member A1"
/notes="match: cDNAs: Em:U24267.1"
join(complement(18284..18345),complement(5834..5927),
complement(5184..5276),complement(2286..2333),
complement(AL954340.4:15217..15372),
complement(AL954340.4:13023..13172),
complement(AL954340.4:12868..12942),
complement(AL954340.4:11444..11631),
complement(AL954340.4:9044..9117),
complement(AL954340.4:7160..7356),
complement(AL954340.4:6952..6999),
complement(AL954340.4:6059..6211),
complement(AL954340.4:5126..5247),
complement(AL954340.4:4207..4325),
complement(AL954340.4:2589..2701))

CDS
      Query Match       90.6%; Score 15.4; DB 5; Length 103216;
      Best Local Similarity 94.1%; Pred. No. 2.2e+03;
      Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCACCCTTGCCCGAG 17
    |||||
Db 43206 GCCCACCCTTGCCCGAG 43190

RESULT 60
AY663406
LOCUS
DEFINITION
  Homo sapiens voucher Coriell Cell Repository DNA sample NA10923 MHC
  class II antigen (HLA-DQB1) gene, HLA-DQB1*0602 allele, MHC
  class II antigen (HLA-DQA1) gene, HLA-DQA1*010201 allele, and
  MHC class II antigen (HLA-DRB1) gene, HLA-DRB1*150101 allele,
  complete cds.
ACCESSION
  AY663406
VERSION
  AY663406.1 GI:52840216
KEYWORDS
  Homo sapiens (human)
SOURCE
  Homo sapiens
  Organism
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
  Homnidae; Homo.
REFERENCE
  1 (bases 1 to 103395)
  Raymond,C.K., Kas,A., Paddock,M., Qiu,R., Zhou,Y., Subramanian,S.,
  Chang,J., Palmieri,A., Haugen,E., Kaul,R. and Olson,M.V.
  Ancient Haplotypes of the HLA Class II region
  Genome Res. 15 (9), 1250-1257 (2005)
  16140993
REFERENCE
  2 (bases 1 to 103395)
  Raymond,C.K., Kas,A., Paddock,M., Qiu,R., Zhou,Y., Subramanian,S.,
  Chang,J., Palmieri,A., Haugen,E., Kaul,R. and Olson,M.V.
  Direct Submission
  Submitted (18-JUN-2004) Genome Center, Department of Medicine,
  University of Washington, Box 352145, Seattle, WA 98195, USA
  Location/Qualifiers
    1..103395
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /specimen_voucher="Coriell Cell Repository DNA sample
    NA10923"
    /db_xref="taxon:9606"
    <13398..>20612
    /gene="HLA-DQB1"
    allele="DQB1*0602"
    join(<14036..14144,15576..15845,18395..18676,19186..19296,
    20394..>20407)
    /gene="HLA-DQB1"
    /allele="DQB1*0602"
    /product="MHC class II antigen"
    join(14036..14144,15576..15845,18395..18676,19186..19296,

gene
mRNA
CDS

20394..20407)
/gene="HLA-DQB1"
/allele="DQB1*0602"
/codon_start=1
/product="MHC class II antigen"
/db_xref="GI:52840219"
/translat="MSWKALRIPGDLRVATVTMLAMLSSLLAEGRSDPEDFVFQFK
GMCYFTNGTERVRLVTRYIYNREEYARFSDVDGVYRAVTPQGRDPAIWNMSQKVLEG
TRAEIDTCVRHNYEVAFGILQRVEPTVITSPSRTEALHNHLLVCSVTDFFPGQIK
VRFRNDQRETAGVSTPLIRNGDWTFOILLMLEMTPORGDVYTCVHEHPSLQSPITV
EWRAQESAKMSLGVGGFVLGLIFLGLLIIRQRSQKGLLH"
complement(37053..43226)
/gene="HLA-DQA1"
/allele="DQA1*010201"
complement(join(<37879..38033,38390..38671,39085..39333,
43103..>43184))
/gene="HLA-DQA1"
/allele="DQA1*010201"
/product="MHC class II antigen"
complement(join(37879..38033,38390..38671,39085..39333,
43103..43184))
/gene="HLA-DQA1"
/allele="DQA1*010201"
/codon_start=1
/product="MHC class II antigen"
/protein_id="AAU88007.1"
/db_xref="GI:52840217"
/translat="MILKALLCALALTTVMSPCGGEDIIVADHVASCGVNLYQFYGP
SGOYTFEFGDEQFYDLERKETAWRPFSKFGGFPQOGLRNMVAKHNLNIMIKR
YNSATANEVPEVTVEFSKSPVTGLQPNLTICLDVNI PFPVNIITLNGSQSVTGVSE
TSFLSKSDHSFFKISYLTFLPSADEIYDKVEHWGLDQPLLKHWEPEIPAPMSELTET
VVCALGSLVGLMGIVGTVFIIQGLRSVGASRHQGPL"
<90424..>101471
/gene="HLA-DRB1"
/allele="DRB1*150101"
join(<90500..90599,95864..96133,98402..98683,99384..99494,
99970..99993,101136..>101149)
/gene="HLA-DRB1"
/allele="DRB1*150101"
/product="MHC class II antigen"
join(90500..90599,95864..96133,98402..98683,99384..99494,
99970..99993,101136..101149)
/gene="HLA-DRB1"
/allele="DRB1*150101"
/codon_start=1
/product="MHC class II antigen"
/protein_id="AAU88008.1"
/db_xref="GI:52840218"
/translat="MVCLKLPGSCMTALTVTLMVLSPLSALSGDTRPRFLWQPKREC
HFNGTERVRLDRIFYNQBESEVRFSDVGEFRVATVTELGPRDAIWNMSQKDIQEARA
AVDTYCRHNYGVVESFTVQRVQPKVTYPSKTPQLQHNLLVCSVSGFYPGSIEVRW
FLNGQESKAGMVTGLIQNGDWTFTLVMLETVPVRSGEVYTCVHEHPSVPLTVIEWR
ARSSAQSKMSLGVGGFVLGLIFLGLFIYFRNQKHSGSLQPTGFLS"

ORIGIN
      Query Match       90.6%; Score 15.4; DB 5; Length 103395;
      Best Local Similarity 94.1%; Pred. No. 2.2e+03;
      Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCACCCTTGCCCGAG 17
    |||||
Db 15871 GCCCACCCTTGCCCGAG 15887

RESULT 61
AY663395
LOCUS
DEFINITION
  Homo sapiens voucher Coriell Cell Repository DNA sample NA14660 MHC
  class II antigen (HLA-DQB1) gene, HLA-DQB1*0602 allele, MHC
  class II antigen (HLA-DQA1) gene, HLA-DQA1*010201 allele, and
  MHC class II antigen (HLA-DRB1) gene, complete cds.
ACCESSION
  AY663395
```

VERSION AY663395.1 GI:52840176  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
REFERENCE 1 (bases 1 to 104996)  
AUTHORS Raymond,C.K., Kas,A., Paddock,M., Qiu,R., Zhou,Y., Subramanian,S.,  
Chang,J., Palmieri,A., Haugen,E., Kaul,R. and Olson,M.V.  
TITLE Ancient haplotypes of the HLA Class II region  
JOURNAL Genome Res. 15 (9), 1250-1257 (2005)  
PUBMED 16140993  
REFERENCE 2 (bases 1 to 104996)  
AUTHORS Raymond,C.K., Kas,A., Paddock,M., Qiu,R., Zhou,Y., Subramanian,S.,  
Chang,J., Palmieri,A., Haugen,E., Kaul,R. and Olson,M.V.  
TITLE Direct Submission  
JOURNAL Submitted (18-JUN-2004) Genome Center, Department of Medicine,  
University of Washington, Box 352145, Seattle, WA 98195, USA  
FEATURES  
source  
1..104996  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/specimen\_voucher="Coriell Cell Repository DNA sample  
NA14660"  
/db\_xref="taxon:9606"  
<2884..>9498  
/gene="HLA-DQB1"  
/alleles="DQB1\*0602"  
join(<2922..3030,4462..4731,7281..7562,8072..8182,  
9280..>9293)  
/gene="HLA-DQB1"  
/alleles="DQB1\*0602"  
/product="MHC class II antigen"  
join(2922..3030,4462..4731,7281..7562,8072..8182,  
9280..>9293)  
/gene="HLA-DQB1"  
/alleles="DQB1\*0602"  
/codon\_start=1  
/product="MHC class II antigen"  
/protein\_id="AAU87980.1"  
/db\_xref="GI:52840179"  
translation="MSWKALRIPGDLRVATVTLMLSLAEGRSDPEDFVFPQK  
GMCVFTNTERVLRLVYIYNREEVAREDSVGVVAVTPQGRPDAYVNSQKVLG  
TRALDVTCHRYEVAFGILQRVEPTVTLSPRTEALHNHLLVCSVTFYFGQIK  
WFRNDOETAGVSTPLIRNGDWTFQILVMLETPRGDVTCHVEHPSLQSPITV  
EWRAQSSAQSKMLSGVGGFVLGFLGLIIRQSRQKGLLH"  
complement(25936..32110)  
/gene="HLA-DQA1"  
/alleles="DQA1\*010201"  
complement(join(<26762..26916,27273..27554,27968..28216,  
31987..>32068))  
/gene="HLA-DQA1"  
/alleles="DQA1\*010201"  
/product="MHC class II antigen"  
complement(join(26762..26916,27273..27554,27968..28216,  
31987..>32068))  
/gene="HLA-DQA1"  
/alleles="DQA1\*010201"  
/codon\_start=1  
/product="MHC class II antigen"  
/protein\_id="AAU87978.1"  
/db\_xref="GI:52840177"  
translation="MILNKALLLGALATTVMSPCGGSDIVADHVASGVNLYFYFGP  
SQYTHFDGEQYDLERKETARWPEFSKFGDFDQGALRNMAVAKHNLMIMKR  
YNSAATNEPDLRVEVFSKSPVTLGQNTLILVDNIFPPVNVITWLSNGQSVTEGVSE  
TSFLSKSDHSPFKISYLTFLPSADEIYDKVHMGLDQPLLKHWPEIPAPMSLLET  
VVCALGLSLVGLMGIWGVTFIILQSLRSGASRHQGL"  
<79388..>90431  
/gene="HLA-DRB1"  
join(<79464..79563,84828..85097,87362..87643,88344..88454,  
88930..88953,90096..>90109)  
/gene="HLA-DRB1"

CDS  
/product="MHC class II antigen"  
join(79464..79563,84828..85097,87362..87643,88344..88454,  
88930..88953,90096..90109)  
/gene="HLA-DRB1"  
/codon\_start=1  
/product="MHC class II antigen"  
/protein\_id="AAU87979.1"  
/db\_xref="GI:52840178"  
translation="MVCLKLPGGSCMTALTVTLMVLSPLALSGDTRPRFLWQPKREC  
HFFNGTERVFLDRHFYQNEESVRFDSVGEFRAVTELGPRDAEYVNSQKDLLEGARA  
AVDTYCRHNYGVESFTVQRVQPKVTYVPSKTPLOHNNLLVCSVGFYPGSIEVRW  
FLNQEEKAGMVTGLIQNGDWTFQILVMLETPRSGEVYTCQVHPSVTSPLTVEWR  
ARSSAQSKMLSGVGGFVLGFLGLIIRQKQKHSGLQPTGFLS"  
ORIGIN  
Query Match 90.6%; Score 15.4; DB 5; Length 104996;  
Best Local Similarity 94.1%; Pred. No. 2.2e+03;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 GCCCACGCTTGGCCGAG 17  
Db 4757 GCCCACCTTGGCCGAG 4773  
RESULT 62  
AE001439\_02/c  
WPCOMMENT  
Sequence split into 17 fragments LOCUS AE001439 Accession AE001439  
Fragment Name Begin End  
AE001439\_00 1 110000  
AE001439\_01 100001 210000  
AE001439\_02 200001 310000  
AE001439\_03 300001 410000  
AE001439\_04 400001 510000  
AE001439\_05 500001 610000  
AE001439\_06 600001 710000  
AE001439\_07 700001 810000  
AE001439\_08 800001 910000  
AE001439\_09 900001 1010000  
AE001439\_10 1000001 1110000  
AE001439\_11 1100001 1210000  
AE001439\_12 1200001 1310000  
AE001439\_13 1300001 1410000  
AE001439\_14 1400001 1510000  
AE001439\_15 1500001 1610000  
AE001439\_16 1600001 1643831  
Continuation (3 of 17) of AE001439 from base 200001 (AE001439 Helicobacter pylori J99, c  
Query Match 90.6%; Score 15.4; DB 15; Length 110000;  
Best Local Similarity 94.1%; Pred. No. 2.2e+03;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 GCCCACGCTTGGCCGAG 17  
Db 109190 GCCCACGCTTGGCCGAG 109174  
RESULT 63  
AE001439\_03/c  
WPCOMMENT  
Sequence split into 17 fragments LOCUS AE001439 Accession AE001439  
Fragment Name Begin End  
AE001439\_00 1 110000  
AE001439\_01 100001 210000  
AE001439\_02 200001 310000  
AE001439\_03 300001 410000  
AE001439\_04 400001 510000  
AE001439\_05 500001 610000  
AE001439\_06 600001 710000  
AE001439\_07 700001 810000  
AE001439\_08 800001 910000  
AE001439\_09 900001 1010000  
AE001439\_10 1000001 1110000

AE001439\_11 1100001 1210000  
AE001439\_12 1200001 1310000  
AE001439\_13 1300001 1410000  
AE001439\_14 1400001 1510000  
AE001439\_15 1500001 1610000  
AE001439\_16 1600001 1643931  
Continuation (4 of 17) of AE001439 from base 300001 (AE001439 Helicobacter pylori J99, c

Query Match 90.6%; Score 15.4; DB 15; Length 110000;  
Best Local Similarity 94.1%; Pred. No. 2.2e+03;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17  
Db 9190 GCCCAGCGTTGGCCGAG 9174

RESULT 64  
CP000240\_27/c  
WPCOMMENT  
Sequence split into 31 fragments LOCUS CP000240 Accession CP000240

Fragment Name	Begin	End
CP000240_00	1	110000
CP000240_01	100001	210000
CP000240_02	200001	310000
CP000240_03	300001	410000
CP000240_04	400001	510000
CP000240_05	500001	610000
CP000240_06	600001	710000
CP000240_07	700001	810000
CP000240_08	800001	910000
CP000240_09	900001	1010000
CP000240_10	1000001	1110000
CP000240_11	1100001	1210000
CP000240_12	1200001	1310000
CP000240_13	1300001	1410000
CP000240_14	1400001	1510000
CP000240_15	1500001	1610000
CP000240_16	1600001	1710000
CP000240_17	1700001	1810000
CP000240_18	1800001	1910000
CP000240_19	1900001	2010000
CP000240_20	2000001	2110000
CP000240_21	2100001	2210000
CP000240_22	2200001	2310000
CP000240_23	2300001	2410000
CP000240_24	2400001	2510000
CP000240_25	2500001	2610000
CP000240_26	2600001	2710000
CP000240_27	2700001	2810000
CP000240_28	2800001	2910000
CP000240_29	2900001	3010000
CP000240_30	3000001	3046682

Continuation (28 of 31) of CP000240 from base 2700001 (CP000240 Cyanobacteria bacterium

Query Match 90.6%; Score 15.4; DB 15; Length 110000;  
Best Local Similarity 94.1%; Pred. No. 2.2e+03;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17  
Db 100616 GCCCAGCGTTGGCCGAG 100600

RESULT 65  
CP000240\_28/c  
WPCOMMENT  
Sequence split into 31 fragments LOCUS CP000240 Accession CP000240

Fragment Name	Begin	End
CP000240_00	1	110000
CP000240_01	100001	210000
CP000240_02	200001	310000
CP000240_03	300001	410000

CP000240\_04 400001 510000  
CP000240\_05 500001 610000  
CP000240\_06 600001 710000  
CP000240\_07 700001 810000  
CP000240\_08 800001 910000  
CP000240\_09 900001 1010000  
CP000240\_10 1000001 1110000  
CP000240\_11 1100001 1210000  
CP000240\_12 1200001 1310000  
CP000240\_13 1300001 1410000  
CP000240\_14 1400001 1510000  
CP000240\_15 1500001 1610000  
CP000240\_16 1600001 1710000  
CP000240\_17 1700001 1810000  
CP000240\_18 1800001 1910000  
CP000240\_19 1900001 2010000  
CP000240\_20 2000001 2110000  
CP000240\_21 2100001 2210000  
CP000240\_22 2200001 2310000  
CP000240\_23 2300001 2410000  
CP000240\_24 2400001 2510000  
CP000240\_25 2500001 2610000  
CP000240\_26 2600001 2710000  
CP000240\_27 2700001 2810000  
CP000240\_28 2800001 2910000  
CP000240\_29 2900001 3010000  
CP000240\_30 3000001 3046682  
Continuation (29 of 31) of CP000240 from base 2800001 (CP000240 Cyanobacteria bacterium

Query Match 90.6%; Score 15.4; DB 15; Length 110000;  
Best Local Similarity 94.1%; Pred. No. 2.2e+03;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17  
Db 616 GCCCAGCGTTGGCCGAG 600

RESULT 66  
CP000248\_28/c  
WPCOMMENT  
Sequence split into 36 fragments LOCUS CP000248 Accession CP000248

Fragment Name	Begin	End
CP000248_00	1	110000
CP000248_01	100001	210000
CP000248_02	200001	310000
CP000248_03	300001	410000
CP000248_04	400001	510000
CP000248_05	500001	610000
CP000248_06	600001	710000
CP000248_07	700001	810000
CP000248_08	800001	910000
CP000248_09	900001	1010000
CP000248_10	1000001	1110000
CP000248_11	1100001	1210000
CP000248_12	1200001	1310000
CP000248_13	1300001	1410000
CP000248_14	1400001	1510000
CP000248_15	1500001	1610000
CP000248_16	1600001	1710000
CP000248_17	1700001	1810000
CP000248_18	1800001	1910000
CP000248_19	1900001	2010000
CP000248_20	2000001	2110000
CP000248_21	2100001	2210000
CP000248_22	2200001	2310000
CP000248_23	2300001	2410000
CP000248_24	2400001	2510000
CP000248_25	2500001	2610000
CP000248_26	2600001	2710000
CP000248_27	2700001	2810000
CP000248_28	2800001	2910000
CP000248_29	2900001	3010000

CP000248\_30 3000001 3110000  
CP000248\_31 3100001 3210000  
CP000248\_32 3200001 3310000  
CP000248\_33 3300001 3410000  
CP000248\_34 3400001 3510000  
CP000248\_35 3500001 3561584  
Continuation (29 of 36) of CP000248 from base 2800001 (CP000248 Novosphingobium aromaticum)

Query Match 90.6%; Score 15.4; DB 15; Length 110000;  
Best Local Similarity 94.1%; Pred. No. 2.2e+03;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCCAGCTTGGCCGAG 17

Db 72861 GCCCAGCTTGGCCGAG 72845

RESULT 67

AE014295\_08/c

WPCOMMENT

Sequence split into 23 fragments LOCUS AE014295 Accession AE014295

Fragment Name Begin End

AE014295\_00 1 110000

AE014295\_01 100001 210000

AE014295\_02 200001 310000

AE014295\_03 300001 410000

AE014295\_04 400001 510000

AE014295\_05 500001 610000

AE014295\_06 600001 710000

AE014295\_07 700001 810000

AE014295\_08 800001 910000

AE014295\_09 900001 1010000

AE014295\_10 1000001 1110000

AE014295\_11 1100001 1210000

AE014295\_12 1200001 1310000

AE014295\_13 1300001 1410000

AE014295\_14 1400001 1510000

AE014295\_15 1500001 1610000

AE014295\_16 1600001 1710000

AE014295\_17 1700001 1810000

AE014295\_18 1800001 1910000

AE014295\_19 1900001 2010000

AE014295\_20 2000001 2110000

AE014295\_21 2100001 2210000

AE014295\_22 2200001 2256640

Continuation (9 of 23) of AE014295 from base 800001 (AE014295 Bifidobacterium longum NC0

Query Match 90.6%; Score 15.4; DB 15; Length 110000;  
Best Local Similarity 94.1%; Pred. No. 2.2e+03;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCCAGCTTGGCCGAG 17

Db 24523 GCCCAGCTTGGCCGAG 24507

RESULT 68

AE017262\_13/c

WPCOMMENT

Sequence split into 29 fragments LOCUS AE017262 Accession AE017262

Fragment Name Begin End

AE017262\_00 1 110000

AE017262\_01 100001 210000

AE017262\_02 200001 310000

AE017262\_03 300001 410000

AE017262\_04 400001 510000

AE017262\_05 500001 610000

AE017262\_06 600001 710000

AE017262\_07 700001 810000

AE017262\_08 800001 910000

AE017262\_09 900001 1010000

AE017262\_10 1000001 1110000

AE017262\_11 1100001 1210000

AE017262\_12 1200001 1310000

AE017262\_13 1300001 1410000

AE017262\_14 1400001 1510000

AE017262\_15 1500001 1610000

AE017262\_16 1600001 1710000

AE017262\_17 1700001 1810000

AE017262\_18 1800001 1910000

AE017262\_19 1900001 2010000

AE017262\_20 2000001 2110000

AE017262\_21 2100001 2210000

AE017262\_22 2200001 2310000

AE017262\_23 2300001 2410000

AE017262\_24 2400001 2510000

AE017262\_25 2500001 2610000

AE017262\_26 2600001 2710000

AE017262\_27 2700001 2810000

AE017262\_28 2800001 2905187

Continuation (14 of 29) of AE017262 from base 1300001 (AE017262 Listeria monocytogenes

Query Match 90.6%; Score 15.4; DB 15; Length 110000;

Best Local Similarity 94.1%; Pred. No. 2.2e+03;

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCCAGCTTGGCCGAG 17

Db 31052 GCCCAGCTTGGCTGAG 31036

RESULT 69

BA000032\_12/c

WPCOMMENT

Sequence split into 19 fragments LOCUS BA000032 Accession BA000032

Fragment Name Begin End

BA000032\_00 1 110000

BA000032\_01 100001 210000

BA000032\_02 200001 310000

BA000032\_03 300001 410000

BA000032\_04 400001 510000

BA000032\_05 500001 610000

BA000032\_06 600001 710000

BA000032\_07 700001 810000

BA000032\_08 800001 910000

BA000032\_09 900001 1010000

BA000032\_10 1000001 1110000

BA000032\_11 1100001 1210000

BA000032\_12 1200001 1310000

BA000032\_13 1300001 1410000

BA000032\_14 1400001 1510000

BA000032\_15 1500001 1610000

BA000032\_16 1600001 1710000

BA000032\_17 1700001 1810000

BA000032\_18 1800001 1877212

Continuation (13 of 19) of BA000032 from base 1200001 (BA000032 Vibrio parahaemolyticus

Query Match 90.6%; Score 15.4; DB 15;

Best Local Similarity 94.1%; Pred. No. 2.2e+03;

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCCAGCTTGGCCGAG 17

Db 33479 GCCCAGCTTGGCCAAG 33463

RESULT 70

AP003159/c

LOCUS ..

DEFINITION Homo sapiens genomic DNA, chromosome lp36 clone:1024F16, complete sequence.

ACCESSION AP003159

VERSION AP003159.1 GI:15208261

KEYWORDS HTG.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

LOCUS 113733 bp DNA linear PRI 16-AUG-2001

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.

#### REFERENCE

1 Machida,T., Ohira,M., Morohashi,A., Mihara,M., Furuta,S., Soeda,E. and Nakagawara,A. 733 genomic DNA of 1p36  
Homo sapiens 113,733 genomic DNA (2001)  
Published Only in Database (2001)  
2 (bases 1 to 113733)  
Machida,T., Ohira,M., Morohashi,A., Mihara,M., Furuta,S., Soeda,E. and Nakagawara,A.

#### TITLE

Direct Submission

#### JOURNAL

Submitted (31-JAN-2001) Akira Nakagawara, Chiba Cancer Center Research Institute, Division of Biochemistry; 666-2 Nitona, Chuoh-ku, Chiba, Chiba 260-8717, Japan  
(E-mail: akiranak@chiba-cc.pref.chiba.jp, Tel:81-43-264-5431, Fax:81-43-265-4459)

#### COMMENT

Additional author information  
Hattori,M., Toyoda,A., Taylor,T.D., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.  
The Institute of Physical and Chemical Research (RIKEN) Genomic Sciences Center (GSC)  
1-7-22 Suehiro-chou, Tsurumi-Ku, Yokohama, Kanagawa, JAPAN zip: 230-0045  
phone: 81-45-503-9111, fax: 81-45-503-9170  
e-mail: hattori@gs.c.riken.go.jp.  
Location/Qualifiers

#### FEATURES

source  
1. 113733  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p36"  
/clone="1024F16"

#### ORIGIN

Query Match 90.6%; Score 15.4; DB 5; Length 113733;  
Best Local Similarity 94.1%; Pred. No. 2.2e+03;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GCCCAGCTTGGCCGAG 17  
|||||  
Db 44326 GCCCAGCTTGGCCGAG 44310

#### RESULT 71

#### AL662789/c

#### LOCUS

DEFINITION Human DNA sequence from clone XXbac-254F23 on chromosome 6 contains the 5' end of the HLA-DRB1 gene for major histocompatibility complex, class II, DR beta 1, the HLA-DQA1 gene for major histocompatibility complex, class II, DQ alpha 1, the HLA-DQB1 gene for major histocompatibility complex, class II, DQ beta 1, a cytochrome C oxidase polypeptide III pseudogene, a major histocompatibility complex, class II, beta polypeptide pseudogene and two CpG islands, complete sequence.

#### ACCESSION

#### VERSION

#### KEYWORDS

AL662789.11 GI:19808179  
HTG; CpG island; cytochrome C oxidase polypeptide III; DQA1; HLA-DQ; HLA-DQA; HLA-DQA1; HLA-DQB; HLA-DQB1; HLA-DRB1; HLA-DRB1; major histocompatibility complex, class II, beta; major histocompatibility complex, class II, DQ alp; major histocompatibility complex, class II, DQ bet; major histocompatibility complex, class II, DR bet.

#### SOURCE

#### ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 147557)  
Tracey,A.

#### REFERENCE

#### TITLE

#### JOURNAL

Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,

#### COMMENT

Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk  
Clone requests: clonerequest@sanger.ac.uk  
On Mar 29, 2002 this sequence version replaced gi:19808178.  
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:  
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep/XXbac-254F23 is from a CHORI-501 human bac - PGF cell line library VECTOR: pTARBAC2.1

This sequence was generated from part of bacterial clone contigs constructed by the MHC Haplotype Consortium and collaborators. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr6/MHC  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: vegas@sanger.ac.uk

#### -----

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.  
Location/Qualifiers

#### source

1. 147557  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/clone="XXbac-254F23"  
/clone\_lib="CHORI-501"

#### misc\_feature

1  
/note="Clone\_left\_end: XXbac-254F23"

#### misc\_feature

6890..7838  
/note="Single clone region. Sequence from reads from a short insert library and a transposon library derived from a single pUC clone. Restriction digest data confirm the assembly."  
44278

#### misc\_feature

/note="Clone\_right\_end: XXbac-161M6"

#### gene

49251..55484

#### mRNA

/gene="HLA-DQA1"  
/locus\_tag="XXbac-BPG254F23.1-001"  
join(49251..49434,53204..53452,53866..54147,54504..54678,54846..55484)  
/gene="HLA-DQA1"

#### CDS

/locus\_tag="XXbac-BPG254F23.1-001"  
/note="match: ESTs: AA381553.1 AV717599.1 BQ049081.1 match: CDNAS: AF037314.1 BC008585.1 M29939.1 M33906.1" join(49353..49434,53204..53452,53866..54147,54504..54658)  
/gene="HLA-DQA1"  
/locus\_tag="XXbac-BPG254F23.1-001"  
/standard\_name="OTTHUMP00000029141"  
/note="match: proteins: O19760 P01909 P04226 Q30410 Q30438"  
/codon\_start=1  
/protein\_id="CAI18230.1"  
/db\_xref="GI:55961973"  
/db\_xref="GOA:Q5V7D5"  
/db\_xref="InterPro:IPR001003"  
/db\_xref="InterPro:IPR003006"  
/db\_xref="InterPro:IPR003597"  
/db\_xref="UniProtKB/TrEMBL:Q5V7D5"  
/translation="MILNKALLGALATTVMSPCGGEDIADVHVASGVNLQYFVGP SGQYTHFDGDEQFYVDLERKETAWRPFESKFGFDPQGLRNWAVAKNINIMIKR YNSTAATNEVPEVTFKSPVTLGPNTLCLVDNIPFPVNVITWSNGQSVTEGVSE TFSLSKSDHSFFKISYILTFPSADEIYDKVEHWGLDQPLKHKWEPEIPAPMSELTT VVCAGLSLGVLGMIVGTVFTIIQGLRSVGSRRHQPL"

```
gene      complement(71364..78711)
          /locus="HLA-DQB1"
          /locus_tag="XXbac-BPG254F23.2-001"
mRNA      complement(join(71364..72143,73241..73351,73861..74142,
76692..76961,78393..78711))
          /gene="HLA-DQB1"
          /locus_tag="XXbac-BPG254F23.2-001"
          /note="match: ESTs: BG542046.1"
          match: cDNAs: M16276.1 M17955.1 M25325.1 M25327.1 M32577.1
X03068.1"
CDS       complement(join(72130..72143,73241..73351,73861..74142,
76692..76961,78393..78501))
          /gene="HLA-DQB1"
          /locus_tag="XXbac-BPG254F23.2-001"
          /standard_name="ORTTHUMP0000029167"
          /note="match: proteins: AAH12106 P01919 Q29884 Q29967
Q30061 Q30091 Q30096"
          /codon_start=1
          /protein_id="CAI18231.1"
          /db_xref="GI:55961974"
          /db_xref="InterPro:IPR000353"
          /db_xref="InterPro:IPR003006"
          /db_xref="InterPro:IPR003597"
          /db_xref="InterPro:IPR007110"
          /translation="MSMKALRIQGLRVATVTLMLSLLAEGRSDPEDFVFQFK
GMCYFTNGTERVLYIYKREVEAFDSVGVRAVTPQGRPAEYVNSQKQVLEG
TRAEIDTCRNVYVARRGILQIRVEPTVTISPSRTEALHNHLLVCSTVDFPQIK
VWFRNDQEEAGVSTPLIRNGDWTFCILVPLEMTFQGRDVTYCHVEHPSLQSPITV
EWRAQSESAQKMLSGVGVGLIFGLIGLIIIRQSRQGLLH"
          complement(118016..118697)
          /locus_tag="XXbac-BPG254F23.4-001"
          /pseudo
          complement(118016..118697)
          /locus_tag="XXbac-BPG254F23.4-001"
          /note="match: proteins: Q99810 Q35486 Q36378 Q92J29 Q94QC1
Q94TJ9 Q94285 Q9G3E4 Q9G3R7"
          /pseudo
          /codon_start=1
          /product="cytochrome c oxidase III pseudogene 1"
          complement(143223..143590)
          /locus_tag="XXbac-BPG254F23.3-001"
          /pseudo
          complement(143223..143590)
          /locus_tag="XXbac-BPG254F23.3-001"
          /note="match: proteins: O19507 O62868 O98263 Q8SPC4 Q95399
Q9BS54"
          /pseudo
          /codon_start=1
          /product="major histocompatibility complex, class II, DQ
beta 3"
          147557
          /note="Clone_right_end: XXbac-254F23"
misc_feature
          147557
ORIGIN
Query Match      90.6%; Score 15.4; DB 5; Length 147557;
Best Local Similarity 94.1%; Pred. No. 2.3e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy      1 GCCCAGCTTGGCCGAG 17
          ||||| ||||| |||||
Db      76666 GCCCACCCTTGGCCGAG 76650

RESULT 72
AC150103/c
LOCUS      AC150103      150372 bp      DNA      linear      HTG 01-JUL-2004
DEFINITION      Gallus gallus clone WAG-102C19, WORKING DRAFT SEQUENCE, 18 ordered
pieces.
ACCESSION      AC150103
VERSION      AC150103.1 GI:49533698
KEYWORDS      HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Gallus gallus (chicken)
ORGANISM      Gallus gallus
```

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus  
1 (bases 1 to 150372)  
DOE Joint Genome Institute.  
Unpublished  
2 (bases 1 to 150372)  
DOE Joint Genome Institute.  
Direct Submission  
Submitted (01-JUL-2004) Production Genomics Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA  
94598-1698, USA  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: <http://www.jgi.doe.gov>  
-----  
Project Information  
Center Project Name: 2390506  
Center clone name: Chk\_102C19  
-----  
Summary Statistics  
Consensus quality: 141126 bases at least Q40  
Consensus quality: 145414 bases at least Q30  
Consensus quality: 147452 bases at least Q20  
Estimated insert size: 171000; agarose-fp estimation  
Estimated insert size: 148672; sum-of-contigs estimation  
Quality coverage: 11.15 in Q20 bases; agarose-fp estimation  
Quality coverage: 12.83 in Q20 bases; sum-of-contigs estimation.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 18 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

COMMENT

1 34543: contig of 34543 bp in length  
\* 34544 34643: gap of unknown length  
\* 34644 38244: contig of 3601 bp in length  
\* 38245 38344: gap of unknown length  
\* 38345 44580: contig of 6236 bp in length  
\* 44581 44680: gap of unknown length  
\* 44681 45294: contig of 4614 bp in length  
\* 45295 49394: gap of unknown length  
\* 49395 59164: contig of 9770 bp in length  
\* 59165 59264: gap of unknown length  
\* 59265 65805: contig of 6541 bp in length  
\* 65806 80351: contig of 14446 bp in length  
\* 80352 80451: gap of unknown length  
\* 80452 83731: contig of 3280 bp in length  
\* 83732 83831: gap of unknown length  
\* 83832 96414: contig of 12583 bp in length  
\* 96415 96514: gap of unknown length  
\* 96515 99192: contig of 2878 bp in length  
\* 99193 100721: contig of 1429 bp in length  
\* 100722 100821: gap of unknown length  
\* 100822 102609: contig of 1788 bp in length  
\* 102610 102709: gap of unknown length  
\* 102710 110148: contig of 7439 bp in length  
\* 110149 110248: gap of unknown length  
\* 110249 116540: contig of 6292 bp in length  
\* 116541 116641: gap of unknown length  
\* 116641 135574: contig of 18934 bp in length  
\* 135575 135675: gap of unknown length  
\* 135675 136613: gap of unknown length  
\* 136614 138347: contig of 1734 bp in length  
\* 138348 138447: gap of unknown length

```

* 138448 150372: contig of 11925 bp in length.
FEATURES
  source      Location/Qualifiers
1. 150372
   /organism="Gallus gallus"
   /mol_type="genomic DNA"
   /db_xref="taxon:9031"
   /clone="WAG-102C19"
   /clone_lib="Texas A&M Wageningen Chicken BAC Library"
34544. 34643
gap          /estimated_length=unknown
38245. 38344
gap          /estimated_length=unknown
44581. 44680
gap          /estimated_length=unknown
49295. 49394
gap          /estimated_length=unknown
59165. 59264
gap          /estimated_length=unknown
65806. 65905
gap          /estimated_length=unknown
80352. 80451
gap          /estimated_length=unknown
83732. 83831
gap          /estimated_length=unknown
96415. 96514
gap          /estimated_length=unknown
99193. 99292
gap          /estimated_length=unknown
100722. 100821
gap          /estimated_length=unknown
102610. 102709
gap          /estimated_length=unknown
110149. 110248
gap          /estimated_length=unknown
116541. 116640
gap          /estimated_length=unknown
135575. 135674
gap          /estimated_length=unknown
136514. 136613
gap          /estimated_length=unknown
138348. 138447
gap          /estimated_length=unknown

ORIGIN
Query Match      90.6%; Score 15.4; DB 12; Length 150372;
Best Local Similarity 94.1%; Pred. No. 2.3e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCTGGCCGAG 17
    |||||
Db 116985 GCCCAGCCTGGCCGAG 116969

RESULT 73
AC169542
LOCUS      AC169542.2 GI:85678605
DEFINITION Bos taurus clone CH240-242B23, WORKING DRAFT SEQUENCE, 9 unordered
           pieces.
ACCESSION  AC169542
VERSION     AC169542.2
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Bos taurus (cattle)
ORGANISM    Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
1 (bases 1 to 150375)
Munry,D., Adams,C., Agbai II,O., Allen,C., Albrooks,S., Archer,P.,
Arredondo,H., Bandaranaike,D., Bangura,L., Beltran,B., Beltran,R.,
Beraducci,A., Biswal,K., Blyth,P., Bonham,H., Buhay,C., Burch,P.,
Cadoree,I., Canada,A., Cardenas,V., Carter,K., Cavazos,I.,
Chacko,J., Chahrouh,M., Chavez,D., Chen,A., Chen,G., Chen,R.,
Cheng,M.-T., Chu,J., Clerc,K., Cockrell,R., Coyle,M., Cree,A.,

```

```

Curry,S., Dai,W., Davila,M.L., Davis,C., Davy-Carroll,L., De
Anda,C., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H.,
Donlin,J., McCauley,S., Dugan-Rocha,S., Dunn,A., Durbin,K.,
Dziuda,D., Egan,A., Escotto,M., Espinosa,V., Eugene,C., Fa,M.,
Fernandez,S., Fernando,P., Flagg,N., Forbes,L., Foster,P.,
Fowler,G., Fu,O., Fuh,E., Garcia,A., Garcia,R., Garner,T.,
Gaskin,C., Gench,S., Ghose,S., Gill,R., Gonzalez,D.,
Gonzalez-Garay,M., Guevara,W., Holder,M., Haaland,W., Haerberlen,K.,
Hall,B., Hamid,H., Hamilton,K., Harbes,B., Harris,R., Haviak,P.,
Hawes,A., Hawkins,E., Hayes,S., Hemphill,L., Hernandez,J.,
Hines,S., Hitchens,M., Hodgson,A., Hogue,M., Hollins,B.,
Howell,L.T., Hulyk,S., Hume,J., Imo,K., Jackson,A., Jackson,L.,
Jacob,L., Jiang,H., Johnson,B., Johnson,R., Kalafus,K., Kelly,S.,
Keys,T., Khan,Z., King,L., Kovar,C., Kowis,A., Kowis,C., Lara,E.,
Leal,S., Lee,K., Lee,S., LeGall,F.I., Lemon,S., Lewis,L., Li,B.,
Li,Y., Li,Z., Linnell,M., Liu,W., Liu,Y.-S., Liu,Y., Liyanage,D.,
London,P., Lopez,J., Lorensuhewa,L., Lozado,R., Luk,T., Madu,R.,
Maheeshwari,M., Mahoney,C., Malloy,K., Mansouri,D., Martinez,E.,
McClelland,H., McPherson,J., Mercadao,C., Metzker,M.,
Milosavljevic,A., Minja,E., Morgan,M., Morris,S., Munidasa,M.,
Murray,D., Nazarith,L., Ngo,D., Nguyen,N., Norwig-Eaughan,E.,
Nott,A., Nwaokemele,O., Obregon,M., Ochi-Okorie,C., Odeh,E.,
Okwuonu,G., Okwuonu,K., Parker,D., Pasternak,S., Patel,B.,
Patel,V., Paul,H., Perez,A., Perez,L., Petrosino,J., Pham,T.,
Primus,E., Pu,L.-L., Puazo,M., Qin,X., Quinn,A., Quiroz,J.,
Rabata,D., Rachlin,E., Reigh,R., Ren,Y., Reuter,M., Richards,S.,
Rives,C., Rodriguez,F., Rojas,A., Ruiz,S.J., Sana,M., Sanders,W.,
Santibanez,J., Santos,R., Savery,G., Scherer,S., Shen,H., Shen,Y.,
Sisson,I., Sneed,A., Sodergren,E., Song,X.-Z., Sorrelle,R.,
Svatek,A., Taylor,E., Taylor,T., Thomas,N., Thorn,R., Thornton,R.,
Trejos,Z., Usmani,K., Vargo,C., Verduzco,D., Villasana,D., Virk,D.,
Voikov,A., Waldron,L., Walker,B., Wang,Q., Wang,S., Warren,J.,
Wei,X., Wheeler,D., Williams,G., Williams,R., Worley,K., Wright,R.,
Wu,J., Yakub,S., Yan,K., Yuan,Y., Yu,F., Zhang,J., Zhang,L.,
Zhang,Z., Zhou,J., Weinstock,G. and Gibbs,R.A.

Direct Submission
Unpublished
2 (bases 1 to 150375)
Worley,K.C.
Direct Submission
Submitted (07-OCT-2005) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 150375)

Bovine Genome Sequencing Consortium
Direct Submission
Submitted (24-JAN-2006) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jan 24, 2006 this sequence version replaced gi:77377088.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: FNFG
Center clone name: CH240-242B23
----- Summary Statistics
Assembly program: Atlas 3.0;

```



Consensus quality: 147151 bases at least Q40  
 Consensus quality: 147915 bases at least Q30  
 Consensus quality: 148615 bases at least Q20  
 Estimated insert size: 153523; sum-of-contigs estimation  
 Quality coverage: 7x in Q20 bases; sum-of-contigs estimation  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 9 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \*  
 \* 1 1995: contig of 1995 bp in length  
 \* 1996 2045: gap of 50 bp  
 \* 2046 33670: contig of 31625 bp in length  
 \* 33671 33720: gap of 50 bp  
 \* 33721 92544: contig of 58824 bp in length  
 \* 92545 92650: gap of 106 bp  
 \* 92651 100413: contig of 7763 bp in length  
 \* 100414 100463: gap of 50 bp  
 \* 100464 103962: contig of 3499 bp in length  
 \* 103963 104012: gap of 50 bp  
 \* 104013 123918: contig of 19906 bp in length  
 \* 123919 123968: gap of 50 bp  
 \* 123969 146272: contig of 22304 bp in length  
 \* 146273 146777: gap of 505 bp  
 \* 146778 148829: contig of 2052 bp in length  
 \* 148830 148929: gap of unknown length  
 \* 148930 150375: contig of 1446 bp in length.  
 \*  
 \* Location/Qualifiers  
 \* 1..150375  
 \* /organism="Bos taurus"  
 \* /mol\_type="genomic DNA"  
 \* /db\_xref="taxon:9913"  
 \* /clone="CH240-242B23"  
 \* 1996..2045  
 \* /estimated\_length=50  
 \* 33671..33720  
 \* /estimated\_length=50  
 \* 92545..92650  
 \* /estimated\_length=106  
 \* 100414..100463  
 \* /estimated\_length=50  
 \* 103963..104012  
 \* /estimated\_length=50  
 \* 123919..123968  
 \* /estimated\_length=50  
 \* 146273..146777  
 \* /estimated\_length=505  
 \* 148830..148929  
 \* /estimated\_length=unknown  
 \*  
 \* ORIGIN  
 \* Query Match 90.6%; Score 15.4; DB 12; Length 150375;  
 \* Best Local Similarity 94.1%; Pred. No. 2.3e+03;  
 \* Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 \*  
 \* QY 1 GCCACGCTTGCCGAG 17  
 \* |||||  
 \* Db 136558 GCCACGCTTGCCGAG 136574  
 \*  
 \* RESULT 74  
 \* AL355273  
 \* LOCUS AL355273 153108 bp DNA linear PRI 18-MAY-2005  
 \* DEFINITION Human DNA sequence from clone RP11-714M16 on chromosome 10 Contains  
 \* the 3' end of the GPRK5 gene for G protein-coupled receptor kinase  
 \* 5 (GRK5) and the RGS10 gene for regulator of G-protein signalling  
 \* 10, complete sequence.

ACCESSION AL355273  
 VERSION AL355273.30 GI:17973918  
 KEYWORDS HTG; GPRK5; GRK5; RGS10.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homidae; Homo.  
 1 (bases 1 to 153108)  
 Kay.M.  
 Direct Submission  
 Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries: [vega@sanger.ac.uk](mailto:vega@sanger.ac.uk)  
 Clone requests: [clonerequest@sanger.ac.uk](mailto:clonerequest@sanger.ac.uk)  
 On Dec 20, 2001 this sequence version replaced gi:17221178.  
 The following abbreviations are used to associate primary accession  
 numbers given in the feature table with their source databases:  
 Em, EMBL; Swi, SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information  
 on the WORMPEP database can be found at  
[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence  
 was generated from part of bacterial clone contigs of human  
 chromosome 10, constructed by the Sanger Centre Chromosome 10  
 Mapping Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/Chr10>  
 RP11-714M16 is from the library RPCI-11.3 constructed by the group  
 of Pieter de Jong. For further details see  
<http://www.chori.org/bacpac/home.htm>  
 VECTOR: pBAC3.6  
 ----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: <http://www.sanger.ac.uk>  
 Contact: [vega@sanger.ac.uk](mailto:vega@sanger.ac.uk)  
 -----  
 This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >=  
 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one subclone; and the assembly was confirmed by restriction digest,  
 except on the rare occasion of the clone being a YAC.  
 FEATURES  
 source  
 1..153108  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="10"  
 /clone="RP11-714M16"  
 /clone\_lib="RPCI-11.3"  
 2000  
 /note="Clone right\_end: RP11-79M19"  
 14785..15226  
 /note="Single clone region. Assembly consistent with  
 restriction digest data."  
 14943  
 /note="Tandem repeat. Forced join. Gap size estimated to  
 be approximately 500bp by restriction digest data."  
 join(AL3555861.19:48518..48898,AL355861.19:167445..167540,  
 AL583824.16:55217..55329,17397..17474,43868..43968,  
 45695..45787,51077..51140,52089..52229,57353..57543,  
 60433..60470,62701..62790,64246..64454,68825..68962,  
 73373..73510,73847..73978,75671..76321)  
 /gene="GPRK5"  
 /locus\_tag="RP11-714M16.1-002"  
 join(AL3555861.19:48518..48898,AL355861.19:167445..167540,  
 AL583824.16:55217..55329,17397..17474,43868..43968,  
 45695..45787,51077..51140,52089..52229,57353..57543,  
 60433..60470,62701..62790,64246..64454,68825..68962,  
 73373..73510,73847..73978,75671..76321)  
 /gene="GPRK5"  
 /locus\_tag="RP11-714M16.1-002"  
 /product="G protein-coupled receptor kinase 5"  
 /note="match: ESTs: AI160744.1 AI338666.1 AI346265.1

```
AI672434.1 AT738477.1 AI830135.1 AU132088.1 AW2999685.1
BF439363.1 BF569124.1 BF972966.1 BG477593.1 BG764871.1
BI232369.1 BI333586.1 BM049553.1 BM562122.1 BM563509.1
BQ029666.1 BQ055527.1 BQ056240.1 BQ065277.1 BQ066445.1
BQ071007.1 BQ184611.1 BQ186585.1 BQ18795.1 BQ327452.1
BQ929071.1 BQ935218.1 BQ936926.1 BU527897.1 BU681200.1
match: cDNAs: LI3388.1"
join(ALJ355861.19:48847. .48898,ALJ355861.19:167445. .167540,
ALJ583824.16:55217. .55329,17397. .17474,43868. .43968,
45695. .45787,51077. .51140,52089. .52229,57353. .57543,
60433. .60470,62701. .62790,64246. .64454,68825. .68962,
73373. .73510,73847. .73978,75671. .75769)
/gene="GPRK5"
/locus tag="RP11-714M16.1-002"
/standard_name="OTTHUMP00000020596"
/notes="match: proteins: P34947 P43249 Q62833 Q8VEB1"
/codon_start=1
/product="G protein-coupled receptor kinase 5"
/protein_id="CAI16487.1"
/db_xref="GI:55958736"
/db_xref="InterPro:IPR000239"
/db_xref="InterPro:IPR000342"
/db_xref="InterPro:IPR000719"
/db_xref="InterPro:IPR008271"
/db_xref="InterPro:IPR011009"
/translation="MELENIIVANTVLLKAREGGGKRGKSKKWEILKPHISQCED
LRRTIDRYGSLCDKPIGRLLRFQCFETPGLEYIQFLDSVAEYVTPDEKLGEKG
KEIMTKYLVTPKPSVFIQAQGDVLVSQTEEKLLQPKCELFSAQSVHEYLGRPFHE
YLDSMFFDRFLQWLKROPVTKNTPROYVLGKGFGVCACOVRAITGKMYACKRLE
KKRIKKRSGSMALNEKILEKUNSOEVNLAYAYETKDALCLVLTIMGDYLKFLY
NMGNPFEEARFYAAEILCGLEDHRENTVYRDLPENILLDDYGHIRISDLGLAV
KIPGEDLIIRVGTVGYMAPEVNNQRYGLSPDYWGIGLCIYEMIEGQSPFRGRKEV
KREEDVRVETVEYVHKFSSEAKSICMLTKDKAQRLGCGEAGAEVKRHPFFRN
MNFKRLAGMLDPPFVDPRAVYCKVDLIEQFSTVGKVLNDTDDDFYKFSFGSVS
IPWQNEIETECRELAVFGNPGLTLPDLNRNHPPEPPKGLLQLRQHQNNKSS
PSSKTSFNHINSHNVSSTGSS"
join(73150. .73510,73847. .73941)
/gene="GPRK5"
/locus tag="RP11-714M16.1-003"
join(73150. .73510,73847. .73941)
/gene="GPRK5"
/locus tag="RP11-714M16.1-003"
/product="G protein-coupled receptor kinase 5"
/notes="match: ESTs: AW170317.1"
75388
/gene="GPRK5"
/locus tag="RP11-714M16.1-002"
/notes="Clone left_end: RP11-128N5"
76298. .76303
/gene="GPRK5"
/locus tag="RP11-714M16.1-002"
76321
/gene="GPRK5"
/locus tag="RP11-714M16.1-002"
/locus tag="RP11-714M16.1-002"
complement(join(120627. .120961,136211. .136354,
146734. .146820,148008. .148179))
/gene="RGS10"
/locus tag="RP11-714M16.2-002"
complement(join(120627. .120961,136211. .136354,
146734. .146820,148008. .148179))
/gene="RGS10"
/locus tag="RP11-714M16.2-002"
/product="regulator of G-protein signalling 10"
/notes="match: ESTs: BQ890645.1 H87415.1
match: cDNAs: AF368902.1"
join(complement(AC012468.10:3196. .3314),
complement(148008. .148126),complement(146734. .146820),
complement(136211. .136354),complement(120627. .120961))
/gene="RGS10"
/locus tag="RP11-714M16.2-001"
join(complement(AC012468.10:3196. .3314),
complement(148008. .148126),complement(146734. .146820),
complement(136211. .136354),complement(120627. .120961))

/gene="RGS10"
/locus tag="RP11-714M16.2-001"
/product="regulator of G-protein signalling 10"
/notes="match: ESTs: BG763741.1 BMD40804.1 BM914660.1
BM914667.1 BQ688292.1 BQ890645.1
match: cDNAs: BC009361.1"
complement(120627)
/gene="RGS10"
/locus tag="RP11-714M16.2-001"
complement(join(120782. .120961,136085. .136157,
146734. .146796))
/gene="RGS10"
/locus tag="RP11-714M16.2-004"
complement(join(120782. .120961,136085. .136157,
146734. .146796))
/gene="RGS10"
/locus tag="RP11-714M16.2-004"
/product="regulator of G-protein signalling 10"
/notes="match: ESTs: BM838412.1"
complement(join(120815. .120961,136211. .136354,
146734. .146820,148008. .148151))
/gene="RGS10"
/locus tag="RP11-714M16.2-002"
/standard_name="OTTHUMP00000020598"
/notes="match: proteins: O43665"
/codon_start=1
/product="regulator of G-protein signalling 10"
/protein_id="CAI16488.1"
/db_xref="GI:55958737"
/translation="MQSELCFADIHSDGSSSSHQSLKSTAKWAASLENLLEDPGV
KRREFUKKFSSENVLFWLACEDFKMQDKTQMQEKAKIYMTFLSKASSQNVVEG
QSLNKKILIEPHLPMFQKLQDQIFNLMKDYSRFLKSLDLFLKHKRTTEEDLPDA
QTAAKRASRIYNT"
join(complement(AC012468.10:3196. .3244),
complement(148008. .148126),complement(146734. .146820),
complement(136211. .136354),complement(120815. .120961))
/gene="RGS10"
/locus tag="RP11-714M16.2-001"
/standard_name="OTTHUMP00000020597"
/notes="match: proteins: Q96GN0"
/codon_start=1
/product="regulator of G-protein signalling 10"
/protein_id="CAI16486.1"
/db_xref="GI:55958735"
/db_xref="GOA:Q96GN0"

Query Match 90.6%; Score 15.4; DB 5; Length 153108;
Best Local Similarity 94.1%; Pred. No. 2.3e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCACCCTTGCCGAG 17
|||||
Db 90763 GCCCACCCTTGCCCAAG 90779

RESULT 75
AC150149/c
LOCUS
DEFINITION Gallus gallus clone WAG-78G23, WORKING DRAFT SEQUENCE, 15 ordered
pieces.
AC150149
AC150149.1 GI:49533744
VERSION HTG: HTGS PHASE2; HTGS DRAFT; HTGS_ACTIVEFIN.
KEYWORDS Gallus gallus (chicken)
SOURCE Gallus gallus
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 160714)
```

AUTHORS DOE Joint Genome Institute.  
TITLE Unpublished  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 160714)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (01-JUL-2004) Production Genomics Facility, DOE Joint Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA 94598-1698, USA

-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: <http://www.jgi.doe.gov>

-----  
Project Information  
Center Project Name: 2381390  
Center clone name: Chk\_78G23

-----  
Summary Statistics  
Consensus quality: 153228 bases at least Q40  
Consensus quality: 156522 bases at least Q30  
Consensus quality: 158225 bases at least Q20  
Estimated insert size: 191000; agarose-fp estimation  
Estimated insert size: 159314; sum-of-contigs estimation  
Quality coverage: 10.76 in Q20 bases; agarose-fp estimation  
Quality coverage: 12.9 in Q20 bases; sum-of-contigs estimation.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 15 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

1 6200: contig of 6200 bp in length  
\* 6201 6300: gap of unknown length  
\* 6301 7878: contig of 1578 bp in length  
\* 7879 7878: gap of unknown length  
\* 7879 8896: contig of 918 bp in length  
\* 8897 8996: gap of unknown length  
\* 8997 13722: contig of 10726 bp in length  
\* 13723 19822: gap of unknown length  
\* 19823 45072: contig of 25250 bp in length  
\* 45073 45172: gap of unknown length  
\* 45173 46466: contig of 1293 bp in length  
\* 46466 46565: gap of unknown length  
\* 46566 48894: contig of 2329 bp in length  
\* 48895 48994: gap of unknown length  
\* 48995 54137: contig of 5143 bp in length  
\* 54138 54237: gap of unknown length  
\* 54238 69612: contig of 15375 bp in length  
\* 69613 69712: gap of unknown length  
\* 69713 82716: contig of 13004 bp in length  
\* 82717 82816: gap of unknown length  
\* 82817 118815: contig of 35999 bp in length  
\* 118816 118915: gap of unknown length  
\* 118916 132678: contig of 13763 bp in length  
\* 132679 132778: gap of unknown length  
\* 132779 139799: contig of 7020 bp in length  
\* 139799 139898: gap of unknown length  
\* 139899 155168: contig of 15270 bp in length  
\* 155169 155268: gap of unknown length  
\* 155269 160714: contig of 5446 bp in length.  
Location/Qualifiers  
1. 160714  
/organism="Gallus gallus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9031"  
/clone\_lib="WAG-78G23"  
6201. .6300  
/estimated\_length=unknown

gap

gap 7879. .7978  
/estimated\_length=unknown  
gap 8897. .8996  
/estimated\_length=unknown  
gap 19723. .19822  
/estimated\_length=unknown  
gap 45073. .45172  
/estimated\_length=unknown  
gap 46466. .46565  
/estimated\_length=unknown  
gap 48895. .48994  
/estimated\_length=unknown  
gap 54138. .54237  
/estimated\_length=unknown  
gap 59613. .69712  
/estimated\_length=unknown  
gap 82717. .82816  
/estimated\_length=unknown  
gap 118816. .118915  
/estimated\_length=unknown  
gap 132679. .132778  
/estimated\_length=unknown  
gap 139799. .139898  
/estimated\_length=unknown  
gap 155169. .155268  
/estimated\_length=unknown

ORIGIN  
Query Match 90.6%; Score 15.4; DB 12; Length 160714;  
Best Local Similarity 94.1%; Pred. No. 2.3e+03;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCCAGCGCTTGGCCGAG 17  
|||||  
Db 26346 GCCCAGCGCTGCGCGAG 26330

RESULT 76  
AC011448  
LOCUS AC011448 165122 bp DNA linear PRI 14-JUL-2002  
DEFINITION Homo sapiens chromosome 19 clone CTC-260F20, complete sequence.  
ACCESSION AC011448  
VERSION AC011448.5 GI:21747436  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
1 (bases 1 to 165122)  
DOE Joint Genome Institute and Stanford Human Genome Center.  
REFERENCE  
AUTHORS  
TITLE Direct Submission  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 165122)  
AUTHORS  
TITLE Direct Submission  
JOURNAL Submitted (07-OCT-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
3 (bases 1 to 165122)  
DOE Joint Genome Institute and Stanford Human Genome Center.  
REFERENCE  
AUTHORS  
TITLE Direct Submission  
JOURNAL Submitted (22-DEC-2000) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
4 (bases 1 to 165122)  
DOE Joint Genome Institute and Stanford Human Genome Center.  
REFERENCE  
AUTHORS  
TITLE Direct Submission  
JOURNAL Submitted (23-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
5 (bases 1 to 165122)  
DOE Joint Genome Institute and Stanford Human Genome Center.  
REFERENCE  
AUTHORS  
TITLE Direct Submission  
JOURNAL Submitted (14-JUL-2002) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

On Jul 14, 2002 this sequence version replaced gi:15281204.  
Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center  
www.shgc.stanford.edu  
Quality: Phrap Quality >=40 99.9% of Sequence;  
Estimated Total Number of Errors is 0.1.

**FEATURES**  
source  
Location/Qualifiers  
1. 165122  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="19"  
/clone="CTC-260F20"

**ORIGIN**  
Query Match 90.6%; Score 15.4; DB 5; Length 165122;  
Best Local Similarity 94.1%; Pred. No. 2.3e+03;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCACGCTGGCGGAG 17  
|||||  
Db 107847 GCCACGCTGGCGGAG 107863

**RESULT 77**  
AC146997  
LOCUS AC146997 168122 bp DNA linear HTG 08-JUL-2004  
DEFINITION Arbacia punctulata clone BAP-144G2, WORKING DRAFT SEQUENCE, 13  
uncloned pieces.  
ACCESSION AC146997  
VERSION AC146997.2 GI:50057837  
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS FULLTOP.  
SOURCE Arbacia punctulata (punctate urchin)  
ORGANISM Arbacia punctulata  
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
Echinoidea; Euechinoidea; Echinacea; Arbacoidea; Arbaciidae;  
Arbacia.

**REFERENCE**  
AUTHORS Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,  
Alsbrooks, S.L., Amaratunga, H.C., Are, J.R., Ayele, M., Banks, T.,  
Barbiera, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D.,  
Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,  
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,  
Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,  
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,  
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,  
Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,  
Delaney, K.J., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,  
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,  
Eamhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,  
Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,  
Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,  
Garrall, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,  
Harris, C., Harris, K., Hart, M., Haviak, P., Hawes, A., Hernandez, J.,  
Hernandez, O., Hodgson, A., Hoques, M., Holloway, C., Hollins, B.,  
Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,  
Jacobson, E., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,  
Karleson, B., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,  
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,  
Li, Z., Li, Z., Licharge, O., Lieu, C., Liu, J., Liu, W., Loulseg, H.,  
Luzado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,  
Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,  
Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,  
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,  
Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,  
Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okwuon, G.,  
Ogungbe, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
Peterson, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,  
Rives, M., Rojas, A., Rojoubkan, I., Rolfe, M., Ruiz, S., Savary, G.,  
Shcher, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I.,  
Sodergren, E., Sonaite, T., Sparks, A., Stanley, H., Stone, H.,  
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
Usmani, K., Vasquez, L., Vera, V., Verduzco, D., Villalon, D.,  
Vinson, R., Waldron, L.M., Wang, Q., Wang, S., Ward-Moore, S.,  
Warren, R., Washington, C., Watlington, S., Williams, G.,  
Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, Y.,  
Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G. and Gibbs, R.

**TITLE**  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 168122)  
AUTHORS Worley, K.C.  
JOURNAL Direct Submission  
TITLE Submitted (31-OCT-2003) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
JOURNAL 3 (bases 1 to 168122)  
AUTHORS Worley, K.C.  
JOURNAL Direct Submission  
TITLE Submitted (08-JUL-2004) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
JOURNAL On Jul 8, 2004 this sequence version replaced gi:38093721.

**COMMENT**  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: http://www.hgsc.bcm.tmc.edu/  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: APAA  
Center clone name: BAP-144G2  
----- Summary Statistics  
Sequencing vector: Plasmid;  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 165376 bases at least Q40  
Consensus quality: 167017 bases at least Q30  
Consensus quality: 168287 bases at least Q20  
Estimated insert size: 178627; sum-of-contigs estimation  
Quality coverage: 17x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 13 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 13948: contig of 13948 bp in length  
\* 13949 14048: gap of unknown length  
\* 14049 32579: contig of 18531 bp in length  
\* 32580 32679: gap of unknown length  
\* 32680 59556: contig of 26877 bp in length  
\* 59557 59656: gap of unknown length  
\* 59657 61694: contig of 2037 bp in length  
\* 61694 61793: gap of unknown length  
\* 61793 63797: contig of 2004 bp in length  
\* 63797 63897: gap of unknown length  
\* 63897 66962: contig of 3065 bp in length  
\* 66963 67062: gap of unknown length  
\* 67063 137626: contig of 70564 bp in length  
\* 137627 137726: gap of unknown length  
\* 137727 141731: contig of 4005 bp in length  
\* 141732 141731: gap of unknown length  
\* 141732 143876: contig of 2045 bp in length  
\* 143877 143976: gap of unknown length  
\* 143977 149300: contig of 5414 bp in length  
\* 149301 149391: gap of unknown length  
\* 149392 151664: contig of 2174 bp in length  
\* 151665 151665: gap of unknown length  
\* 151666 159005: contig of 7241 bp in length

\* 159006 159105: gap of unknown length  
 \* 159106 168122: contig of 9017 bp in length.  
 FEATURES  
 source

1. .168122  
 /organism="Arbacia punctulata"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:7641"  
 /clone="BAP-144G2"  
 13949 .14048  
 /estimated\_length=unknown  
 32580 .32679  
 /estimated\_length=unknown  
 59557 .59656  
 /estimated\_length=unknown  
 61694 .61793  
 /estimated\_length=unknown  
 63798 .63897  
 /estimated\_length=unknown  
 66963 .67062  
 /estimated\_length=unknown  
 137627 .137726  
 /estimated\_length=unknown  
 141732 .141831  
 /estimated\_length=unknown  
 143877 .143976  
 /estimated\_length=unknown  
 149391 .149490  
 /estimated\_length=unknown  
 151665 .151764  
 /estimated\_length=unknown  
 159006 .159105  
 /estimated\_length=unknown

## ORIGIN

Query Match 90.6%; Score 15.4; DB 12;  
 Best Local Similarity 94.1%; Pred. No. 2.3e+03;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGGCCGAG 17  
 | ||||| ||||| |||||  
 Db 80398 GACCACGCTTGGCCGAG 80414

## RESULT 78

AC130666/c  
 LOCUS AC130666 172995 bp DNA linear ROD 29-SEP-2004  
 DEFINITION Mus musculus chromosome 5, clone RP24-7907, complete sequence.  
 ACCESSION AC130666  
 VERSION AC130666.9 GI:52839751  
 KEYWORDS HTG.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidae; Muridae; Murinae; Mus.

## REFERENCE

AUTHORS Birren,B., Nusbaum,C. and Lander,E.  
 TITLE Mus musculus chromosome 5, clone RP24-7907  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 172995)  
 AUTHORS Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,

Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Charazou,B., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faros,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,

Roman,J., Roy,A., Schauer,S., Schupack,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

Submitted (13-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 172995)  
 AUTHORS Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B., Dearellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faros,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hagos,B., Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R., MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupack,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

## Direct Submission

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

Submitted (06-AUG-2004) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 4 (bases 1 to 172995)

AUTHORS Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B., Dearellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faros,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hagos,B., Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R., MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupack,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

## Direct Submission

## TITLE

## JOURNAL

## COMMENT

Submitted (29-SEP-2004) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Sep 29, 2004 this sequence version replaced gi:50511630.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence submissions@broad.mit.edu

----- Project Information

Center project name: L27060

Center clone name: 79\_O\_7

-----

## FEATURES

## source

Location/Qualifiers

1. .172995



Web site: <http://www.sanger.ac.uk>  
 Contact: [vega@sanger.ac.uk](mailto:vega@sanger.ac.uk)

-----  
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

## FEATURES

## source

## Location/Qualifiers

```

1..179959
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="13"
/map="q33.1-34"
/clone="RP11-474D23"
/clone_lib="RPCI-11.2"

1
/misc_feature
  /note="Clone left end: RP11-474D23"
  join(4130..4224,4319..4410,11645..11755,23325..23394,26218..26393,27545..27709)
  /gene="C13orf16"
  /locus_tag="RP11-474D23.1-002"
  join(4130..4224,4319..4410,11645..11755,23325..23394,26218..26393,27545..27709)
  /gene="C13orf16"
  /locus_tag="RP11-474D23.1-002"
  /product="chromosome 13 open reading frame 16"
  /note="match: ESTs: A1004422.1 AW070313.1 AW665078.1 BF509929.1 B1828648.1 B1830521.1 match: cDNA8: BC029889.1"
  join(413086..1284885..84958,4319..4410,11645..11755,22756..22900,23325..23394,26218..26393,27545..27711)
  /gene="C13orf16"
  /locus_tag="RP11-474D23.1-001"
  join(413086..1284885..84958,4319..4410,11645..11755,22756..22900,23325..23394,26218..26393,27545..27711)
  /gene="C13orf16"
  /locus_tag="RP11-474D23.1-001"
  /product="chromosome 13 open reading frame 16"
  /note="match: ESTs: AA369808.1 AA470135.1 AW957098.1"
  join(413086..1284912..84958,4319..4410,11645..11755,22756..22760)
  /gene="C13orf16"
  /locus_tag="RP11-474D23.1-001"
  /standard_name="OTTHUMP0000018719"
  /codon_start=1
  /product="chromosome 13 open reading frame 16"
  /protein_id="CAH74151.1"
  /db_xref="GI:55662217"
  /db_xref="InterPro:IPR000519"
  /db_xref="UniProtKB/TrEMBL:Q5VZG8"
  /translation="MPATCGQGVTVQACARPLHLSAAAMEVLEVKNSPRHLKQ FTVCDVPLYD(CDYNVSRDR)CQELGCGCEGVCYKAVPI"
  join(4353..4410,11645..11755,23325..23394,26218..26393,27545..27585)
  /gene="C13orf16"
  /locus_tag="RP11-474D23.1-002"
  /standard_name="OTTHUMP0000018720"
  /codon_start=1
  /product="chromosome 13 open reading frame 16"
  /protein_id="CAH74150.1"
  /db_xref="GI:55662216"
  /db_xref="InterPro:IPR000519"
  /db_xref="UniProtKB/TrEMBL:Q8N6K0"
  /translation="WEVLEVKNSPRHLKQFTVCDVPLYD(CDYNVSRDR)CQELGCGCEGVCYKAVPIIHFSAALIVIIAGAFVITIIYRVIQESRKEKALPDVVALPKQSS EKAELASSSKLGLKGPASPGPPSAGPSNKSDEKDDVTGTTTEAETED"
  join(4304..14058)
  /gene="C13orf16"
  /locus_tag="RP11-474D23.1-001"

```

## polyA\_signal

/note="Sequence from overlapping clone bA494P5 (AL139086). Assembly confirmed by restriction digest."

27889..27694

/gene="C13orf16"

/locus\_tag="RP11-474D23.1-001"

## polyA\_site

27709

/gene="C13orf16"

/locus\_tag="RP11-474D23.1-001"

## misc\_feature

37339

/note="Clone right end: RP11-494P5"

40640..41524

/note="Single clone region. Sequence from reads from a short insert library derived from a clone PCR. Restriction digest data confirm the assembly."

## misc\_feature

41692..41696

/note="Single clone region. Sequence from reads from a short insert library derived from a clone PCR. Restriction digest data confirm the assembly."

## misc\_feature

59653..59994

/note="Single clone region. Sequence from reads from a short insert library derived from a clone PCR. Restriction digest data confirm the assembly."

## misc\_feature

118600

/note="Tandem repeat. Forced join. Gap size estimated to be approximately 500bp by restriction digest data."

## misc\_feature

17889..178929

/note="Sequence from overlapping clone bA190K22 (AL162714). Assembly confirmed by restriction digest data."

## misc\_feature

179959

/note="Clone right end: RP11-474D23"

## ORIGIN

Query Match 90.6%; Score 15.4; DB 5; Length 179959;  
 Best Local Similarity 94.1%; Pred. No. 2.3e+03;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17  
 |||||  
 Db 145314 GCCCAGCGTTGGCCGAG 145298

## RESULT 80

AC017033/c AC017033 183887 bp DNA linear PRI 21-APR-2005  
 LOCUS Homo sapiens BAC clone RP11-297J22 from 2, complete sequence.  
 DEFINITION  
 ACCESSION AC017033  
 VERSION AC017033.5 GI:12740494  
 KEYWORDS HTG.

## SOURCE

ORGANISM Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homidae; Homo

## REFERENCE

1 (bases 1 to 183887)  
 Maupin,R., Kruchowski,S. and Haakenson,W.  
 The sequence of Homo sapiens BAC clone RP11-297J22  
 Unpublished (2001)

## JOURNAL

2 (bases 1 to 183887)  
 Waterston,R.H.  
 Direct Submission  
 Submitted (09-DEC-1999) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA

## AUTHORS

3 (bases 1 to 183887)  
 Waterston,R.H.  
 Direct Submission  
 Submitted (10-FEB-2001) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA

## TITLE

4 (bases 1 to 183887)

```

AUTHORS
TITLE Direct Submission
JOURNAL Submitted (11-FEB-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE
AUTHORS 5 (bases 1 to 183887)
TITLE Waterston,R.
JOURNAL Direct Submission
Submitted (09-MAY-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE
AUTHORS 6 (bases 1 to 183887)
TITLE Waterston,R.
JOURNAL Direct Submission
Submitted (10-MAY-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE
AUTHORS 7 (bases 1 to 183887)
TITLE Wilson,R.K.
JOURNAL Direct Submission
Submitted (21-APR-2005) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT ----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@watson.wustl.edu
----- Summary Statistics
-----
Center project name: H_NH0297J22
-----
NOTICE:
This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.
MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren,
Department of Genetics, Washington University, St. Louis MO. For
additional information about the map position of this sequence, see
http://genome.wustl.edu
SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male
donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
and coworkers at http://www.chori.org
VECTOR: pBACe3.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-57J18; the clone sequenced
to the right is RP11-549G13. Actual start of this clone is at base
position 1 of RP11-297J22; actual end is at base position 183887 of
RP11-297J22.
RP11-297J22 contains a tandem repeat between bases 149996 to 150321
where there are low quality bases. Digest and pcr information
confirm the assembly.
There are polymorphisms between RP11-297J22 and RP11-549G13.
Location/Qualifiers
1..183887
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
FEATURES
source
/chromosome="2"
/clone="RP11-297J22"
/clone_lib="RPCI-11"
3664..6040
/misc_feature
/note="CpG island (GC=67.4, o/e=0.81, #CpGs=212)"
9984..10384
/misc_feature
/note="CpG island (GC=77.1, o/e=1.02, #CpGs=61)"
60665..65850
/gene
/gene="GLI2"
join(60665..60820,65673..65850)
/gene="GLI2"
/product="unknown"
/note="Homo sapiens GLI-Kruppel family member GLI2 (GLI2),
transcript variant 1, mRNA.; H_NH0297J22.1
This gene was based on gi(135182230)
Continues as H_NH0549G13.1"
135634..136015
/misc_feature
/note="CpG island (GC=68.6, o/e=0.67, #CpGs=34)"
ORIGIN
Query Match 90.6%; Score 15.4; DB 5; Length 183887;
Best Local Similarity 94.1%; Pred. No. 2.3e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 GCCACGCTGGCGGAG 17
Db 4738 GCCGCGCTGGCGGAG 4722
RESULT 81
AC019078
LOCUS AC019078 187027 bp DNA linear HTG 30-NOV-2000
DEFINITION Homo sapiens chromosome 2 clone RP11-330P9, WORKING DRAFT SEQUENCE,
22 unordered pieces.
ACCESSION AC019078
VERSION AC019078.6 GI:11465145
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.
REFERENCE
1 (bases 1 to 187027)
Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 187027)
Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (30-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Nov 30, 2000 this sequence version replaced gi:9887808.
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0330P09
----- Summary Statistics -----
Sequencing vector: M13; 99%
Sequencing vector: plasmid; 0%
Chemistry: Dye-terminator ET; 99% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 171161 bases at least Q40
Consensus quality: 176449 bases at least Q30
Consensus quality: 179564 bases at least Q20
Insert size: 180000; agarose-fp
Insert size: 184927; sum-of-ctngs
Quality coverage: 4.11 in Q20 bases; agarose-fp
Quality coverage: 4.12 in Q20 bases; sum-of-ctngs

```



-----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 22 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

```

* 1      1119: contig of 1119 bp in length
* 1120   1219: gap of unknown length
* 1220   2642: contig of 1423 bp in length
* 2643   2742: gap of unknown length
* 2743   7458: contig of 4716 bp in length
* 7459   7558: gap of unknown length
* 7559   11010: contig of 3452 bp in length
* 11011  11110: gap of unknown length
* 11111  15060: contig of 3950 bp in length
* 15061  15160: gap of unknown length
* 15161  19635: contig of 4475 bp in length
* 19636  19735: gap of unknown length
* 19736  24016: contig of 4281 bp in length
* 24017  24116: gap of unknown length
* 24117  29667: contig of 5551 bp in length
* 29668  29767: gap of unknown length
* 29768  35444: contig of 5677 bp in length
* 35445  35544: gap of unknown length
* 35545  40954: contig of 5410 bp in length
* 40955  48233: contig of 7179 bp in length
* 48234  48334: gap of unknown length
* 48334  57406: contig of 9073 bp in length
* 57407  57506: gap of unknown length
* 57507  66359: contig of 8853 bp in length
* 66360  66460: gap of unknown length
* 66460  76815: contig of 10355 bp in length
* 76815  76915: gap of unknown length
* 76915  87795: contig of 10781 bp in length
* 87795  98686: gap of unknown length
* 98687  98786: contig of 10891 bp in length
* 98787  112322: contig of 13536 bp in length
* 112323 112423: gap of unknown length
* 112423 126002: contig of 13580 bp in length
* 126003 126102: gap of unknown length
* 126103 139276: contig of 13174 bp in length
* 139277 139376: gap of unknown length
* 139377 159274: contig of 19898 bp in length
* 159275 159375: gap of unknown length
* 159375 185651: contig of 26277 bp in length
* 185652 185751: gap of unknown length
* 185752 187027: contig of 1276 bp in length.
```

## FEATURES

```

source
1..187027
  /organism="Homo sapiens"
  /mol_type="genomic DNA"
  /db_xref="taxon:9606"
  /chromosome="2"
  /clone="RP11-330P9"
misc_feature
1..1119
  /note="assembly_name:Contig10"
gap
1120..1219
  /estimated_length=unknown
misc_feature
1220..2642
  /note="assembly_name:Contig13"
gap
2643..2742
  /estimated_length=unknown
misc_feature
2743..7458
  /note="assembly_name:Contig14"
gap
7459..7558
  /estimated_length=unknown
misc_feature
7559..11010
  /note="assembly_name:Contig15"
```

```

gap
11011..11110
  /estimated_length=unknown
misc_feature
11111..15060
  /note="assembly_name:Contig16"
gap
15061..15160
  /estimated_length=unknown
misc_feature
15161..19635
  /note="assembly_name:Contig17"
gap
19636..19735
  /estimated_length=unknown
misc_feature
19736..24016
  /note="assembly_name:Contig18"
gap
24017..24116
  /estimated_length=unknown
misc_feature
24117..29667
  /note="assembly_name:Contig19"
gap
29668..29767
  /estimated_length=unknown
misc_feature
29768..35444
  /note="assembly_name:Contig20"
gap
35445..35544
  /estimated_length=unknown
misc_feature
35545..40954
  /note="assembly_name:Contig21"
gap
40955..41054
  /estimated_length=unknown
misc_feature
41055..48233
  /note="assembly_name:Contig22"
gap
48234..48333
  /estimated_length=unknown
misc_feature
48334..57406
  /note="assembly_name:Contig23"
gap
57407..57506
  /estimated_length=unknown
misc_feature
57507..66359
  /note="assembly_name:Contig24"
gap
66360..66459
  /estimated_length=unknown
misc_feature
66460..76814
  /note="assembly_name:Contig25"
gap
76815..76914
  /estimated_length=unknown
misc_feature
76915..87795
  /note="assembly_name:Contig26"
gap
87696..87795
  /estimated_length=unknown
misc_feature
87796..98686
  /note="assembly_name:Contig27"
gap
98687..98786
  /estimated_length=unknown
misc_feature
98787..112322
  /note="assembly_name:Contig28"
gap
112323..112422
  /estimated_length=unknown
misc_feature
112423..126002
  /note="assembly_name:Contig29"
gap
126003..126102
  /estimated_length=unknown
misc_feature
126103..139276
  /note="assembly_name:Contig30"
gap
139277..139376
  /estimated_length=unknown
misc_feature
139377..159274
  /note="assembly_name:Contig31"
gap
159275..159374
  /estimated_length=unknown
misc_feature
159375..185651
  /note="assembly_name:Contig32"
gap
185652..185751
  /estimated_length=unknown
misc_feature
185752..187027
  /note="assembly_name:Contig3"
```

ORIGIN

```
Query Match      90.6%; Score 15.4; DB 12; Length 187027;
Best Local Similarity 94.1%; Pred. No. 2.3e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17
Db 154552 GCCCGCGCTTGGCCGAG 154568

RESULT 82
AC150074/c
LOCUS      AC150074      193015 bp      DNA      linear      HTG 01-JUL-2004
DEFINITION Gallus gallus clone CH261-96B15, WORKING DRAFT SEQUENCE, 9 ordered
            pieces.
ACCESSION   AC150074
VERSION     AC150074.1 GI:49533669
KEYWORDS    HTG; HTGS PHASE2; HTGS DRAFT; HTGS_ACTIVEFIN.
SOURCE      Gallus gallus (chicken)
ORGANISM    Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1 (bases 1 to 193015)
AUTHORS     DOE Joint Genome Institute.
TITLE       Unpublished
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 193015)
AUTHORS     DOE Joint Genome Institute.
TITLE       Direct Submission.
JOURNAL     Submitted (01-JUL-2004) Production Genomics Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA
            94598-1698, USA
COMMENT     -----Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -----
Project Information
Center Project Name: 2967678
Center clone name: JF2-96B15
-----
Summary Statistics
Consensus quality: 189480 bases at least Q40
Consensus quality: 190386 bases at least Q30
Consensus quality: 191284 bases at least Q20
Estimated insert size: 204000; agarose-fp estimation
Estimated insert size: 192215; sum-of-contigs estimation
Quality coverage: 9.75 in Q20 bases; agarose-fp estimation
Quality coverage: 10.35 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
-----
1 74910: contig of 74910 bp in length
* 74911 75010: gap of unknown length
* 75011 79172: contig of 4162 bp in length
* 79173 79272: gap of unknown length
* 79273 89913: contig of 10641 bp in length
* 89914 90013: gap of unknown length
* 90014 104546: contig of 14533 bp in length
* 104547 104646: gap of unknown length
* 104647 112802: contig of 8156 bp in length
* 112803 112902: gap of unknown length
* 112903 130534: contig of 17632 bp in length
* 130535 130634: gap of unknown length
* 130635 133851: contig of 3217 bp in length
* 133852 133951: gap of unknown length
```

```
* 133952 136970: contig of 3019 bp in length
* 136971 137070: gap of unknown length
* 137071 193015: contig of 55945 bp in length.
FEATURES             source
     1..193015
     /organism="Gallus gallus"
     /mol_type="genomic DNA"
     /db_xref="taxon:9031"
     /clone_lib="CH261-96B15"
     74911..75010
     /estimated_length=unknown
     79173..79272
     /estimated_length=unknown
     89914..90013
     /estimated_length=unknown
     104547..104646
     /estimated_length=unknown
     112803..112902
     /estimated_length=unknown
     130535..130634
     /estimated_length=unknown
     133852..133951
     /estimated_length=unknown
     136971..137070
     /estimated_length=unknown

ORIGIN
Query Match      90.6%; Score 15.4; DB 12; Length 193015;
Best Local Similarity 94.1%; Pred. No. 2.3e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17
Db 111419 GCCCAGCGTTGGCCGAG 111403

RESULT 83
AX417035
LOCUS      AX417035      195269 bp      DNA      linear      PAT 02-SEP-2002
DEFINITION Sequence 4026 from Patent WO0228891.
ACCESSION   AX417035
VERSION     AX417035.1 GI:21449645
KEYWORDS    Listeria innocua
SOURCE      Listeria innocua
ORGANISM    Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
REFERENCE   1
AUTHORS     Kunst, F. and Glaser, P.
TITLE       Listeria innocua genome and applications
JOURNAL     Patent: WO 0228891-A 4026 11-APR-2002;
            INSTITUT PASTEUR (FR) ; CENTRE NATIONAL DE LA RECHERCHE
            SCIENTIFIQUE (CNRS) (FR)
FEATURES             source
     1..195269
     /organism="Listeria innocua"
     /mol_type="unassigned DNA"
     /db_xref="taxon:1642"
     /note="second part of seq 8
     first part = seq 8 (000001 - 349980)
     second part = seq 4026 (300001 - 495269)"

ORIGIN
Query Match      90.6%; Score 15.4; DB 2; Length 195269;
Best Local Similarity 94.1%; Pred. No. 2.3e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17
Db 100813 GCCCAGCGTTGGCTGAG 100829

RESULT 84
```

```

AC002365/c
LOCUS       AC002365               196542 bp    DNA    linear    PRI 22-AUG-1998
DEFINITION  Homo sapiens chromosome X clone U177G4, U152H5, U168D5, 174A6,
U172D6, and U186B3 from Xp22, complete sequence.
ACCESSION   AC002365 U82630 U83111
VERSION     AC002365.1 GI:2358015
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
1 (bases 1 to 196542)
Muzny,D., Aronson,A.D., Adams,C., Brundage,E., Bunac,C.,
Carvelli,K., Chacko,J., Chen,J., Di,W., Ding,Y., Dugan,S.,
Durbin,J., Forcum,J., Ganesh,R., Garcia,C., Goodman,M.,
Correll,J.H., Haywood,M., Hernandez,J., Jackson,L., Jin,S.,
Kampal,R., Karpthy,S., Kovar,C., Leal,B., Li,Y., Lichtarge,O.,
Liu,W., Logan,O., Lu,J., Ly,T., Martinez,C., Oswal,G., Perez,L.,
Rashid,N.D., Rowland,K., Savage,L., Scherer,S.E., Shen,H.,
Simon,M., Stovall,K., Timms,K.M., Todd,J., Vo,Q., Williamson,A.,
Worley,K.C., Yu,W., Chinault,C., Nelson,D. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 196542)
Chiu,M.W.
Direct Submission
Submitted (23-JUL-1997) Molecular and Human Genetics, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 196542)
Muzny,D., Ansari-Lari,M.A., Timms,K.M., Yu,W., Dugan,S., Lu,J.,
Shen,Y., Rowland,K., Liu,W., Perez,L., Ding,Y., Gonzalez,O.,
Haywood,M., Jain,A., Leal,B., Logan,O., Nguyen,V., Savage,L.,
Shen,H., Worley,K., Chen,E., Forcum,J., Aronson,A.D., Chiu,M.W.,
Correll,J.H., Brundage,E., Di,W., Chinault,C., Nelson,D. and
Gibbs,R.A.
Direct Submission
Submitted (20-AUG-1997) Molecular and Human Genetics, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 196542)
Chiu,M.W.
Direct Submission
Submitted (04-SEP-1997) Molecular and Human Genetics, Baylor
College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA
5 (bases 1 to 196542)
Worley,K.C.
Direct Submission
Submitted (22-AUG-1998) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On or before Sep 4, 1997 this sequence version replaced gi:1766071,
gi:177380, gi:2289024.
Sequencing is completed to a minimum standard of double strand
coverage with a minimum of 2 clones and 2 reads with no ambiguities
or 2 chemistries with a minimum of 2 clones and 3 reads with no
ambiguities. If the sequence quality does not meet this standard,
it will be indicated in the annotation.
The repeat regions shown were identified using RepeatMasker by
Adrian Smit.
Sequence similarities were identified using Powerblast by Jinghui
Zhang.
Exon/Intron boundaries of identified genes were chosen if there
were canonical splice junctions that maintained sequence continuity
across the splice junctions.
Location/Qualifiers
1. 196542
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="X"
FEATURES             source
repeat_region       /map="Xp22"
repeat_region       /clone="U177G4"
repeat_region       /cclone_lib="Livermore Chromosome X Cosmid Library
LLOXNC01"
repeat_region       complement(760..1057)
repeat_region       /rpt_family="AluJb"
repeat_region       1893..2194
repeat_region       /rpt_family="AluSg"
repeat_region       2305..2584
repeat_region       /rpt_family="AluSx"
repeat_region       3418..3630
repeat_region       /rpt_family="MER20"
repeat_region       complement(3806..4096)
repeat_region       /rpt_family="AluSx"
repeat_region       4138..4537
repeat_region       /rpt_family="L1MB2"
repeat_region       complement(4549..4841)
repeat_region       /rpt_family="AluJb"
repeat_region       complement(5421..5528)
repeat_region       /rpt_family="MER58B"
repeat_region       complement(7001..7132)
repeat_region       /rpt_family="FAM"
repeat_region       7773..7903
repeat_region       /rpt_family="AluJ"
repeat_region       7910..8215
repeat_region       /rpt_family="AluYa5"
repeat_region       8224..8401
repeat_region       /rpt_family="AluJ"
repeat_region       complement(8932..9232)
repeat_region       /rpt_family="AluY"
repeat_region       complement(9458..9481)
repeat_region       /rpt_family="POLY_A"
repeat_region       11579..11618
repeat_region       /rpt_family="AT rich"
repeat_region       complement(11619..11916)
repeat_region       /rpt_family="AluSg"
repeat_region       complement(11917..12252)
repeat_region       /rpt_family="AluJb"
repeat_region       complement(12328..12633)
repeat_region       /rpt_family="AluSx"
repeat_region       complement(12665..13117)
repeat_region       /rpt_family="L1MB3"
repeat_region       complement(13236..13267)
repeat_region       /rpt_family="L1MA1"
repeat_region       13323..13604
repeat_region       /rpt_family="AluJb"
misc_feature        13983..14378
misc_feature        /note="Similar to sequence with GenBank Accession number
Z36840"
repeat_region       15384..15565
repeat_region       /rpt_family="MER58B"
repeat_region       15656..15956
repeat_region       /rpt_family="AluSg"
repeat_region       16091..16660
repeat_region       /rpt_family="L1MB7"
gene               <17023..>92814
gene               /gene="HSAPXL"
gene               join(<17023..17176,34350..34481,37731..40071,41563..41663,
75348..76243,80507..81058,82568..82739,88014..88286,
90044..>92814)
gene               /gene="HSAPXL"
mrna               join(<17023..17176,34350..34481,37731..40071,41563..41663,
75348..76243,80507..81058,82568..82739,88014..88286,
90044..>92814)
mrna               /gene="HSAPXL"
CDS               join(<17023..17176,34350..34481,37731..40071,41563..41663,
75348..76243,80507..81058,82568..82739,88014..88286,
90044..90310)
CDS               /gene="HSAPXL"
CDS               /note="human homolog of Xenopus laevis APX gene, similar
to sequence encoded by GenBank Accession Number X83543"
CDS               /codon_start=3

```

```
/product="APXL"
/protein_id="AAC32592.1"
/db_xref="GI:11773381"
/translation="IEGSKAAAVDKLLAGDEIVGINDIGLSGFPROBAICLVKGSHTK
LKLVKRSRSLQRLTHFSSLSGSDSLDHPSELAKSPFTSTSGCPSMSGRHSHASSSHDL
SSWEQNTLQRTLDHSLDHPSSRLSVAKNSSIDHLGSHSKRDSAYGSEST
SSSTPDHLKADTSSAENILYTVGLWEAPRQGGQAAAGDPQGSSEKUSCFPRVP
GDSGGRPRYNAEPKLAAPGRNFGVWYVDPKKAPSPPPPPPLRSDSFAATKS
HEKAQGVFSEAAAAAFTALAAQAPGRDRPELTDPMRSAPHSGLSGSGGPGCPQ
EAHADGSPKSDGARSQSLSSSDVRFQPSHGRHPLLYSDHSPCADSLGQBP
GAASFONDSPOVGLSSCDQKLGSGWQGRPCVQDQLAAQLWAGCWPDSLTGALE
SLPPTVQGPRLHLPQEGPPDARETRGRCYPLDKGAGGCSAGAQEPPRASRAEKASQ
RLAASITWAGESSRIQCPETPLLSLUTQEGKRPESSPDSATPPPPFPAHVCKPTR
RSRFAITLNEITSMRAIKQKSTALTAAGEADGTGRWAGLGGGTQEGPLACT
YKDLKBAQARVLRATFEKDLDPNFDLYPESLEHMGDPDPTVPHFWEAGLAQPS
STSGGPHPRIGRRRTAEQKLKSYSEPEKMEVGLTRGYSQHQPRTSEDTVTGFA
DRMKFEETSKVPQAPQALHGIIPDKPERPRTAGTCGTEPMSRTTSLGDSIN
AHSNAEAGTSDFPRLGTAEVQASWKEQRKPLEARSSGRCHSADDILOVSLDPQBR
FQVHGRSSPSTDIHYQASVELRQAGDPGEFBEELPSAVRAEEGQSTPQADAO
CREGSPGQHPSPSKAPNPPTFSELSCRGAPELPREGRAGTLPRDYRISEESTP
ADLGPRAQSPGSLHARGQDSWPVSSALLSKRPAPQPPPPKREPRYRATDGAPADA
PVGVLGRPFTTSPASLDVVARLSLHSPSVFSSAQPDTPKATVCERGSHVSGDA
SRPLPEALLPQKQHLBLQATMTETSSPSPQAPAKLTDKPLLIODEDSTRIERVM
DNNTVXMKVIKIVHSSQPEKESRQSLACAPPEPALPHGLEKDQIKTLSTSQFYSR
FCLYTRQAGSEAPRAQPAEPQPLGTQVPEKDRCTSPGLSTYMKAKEKTVEDLKS
ELAREIVGDKSLADILDPVKIKITMDLMEGIPDKDEHLLLEEAQRRKLLPKTIPSR
STEERKEEPSVPAALVATNSTYTSAPKAELLIKNKDLQEQHEEDSGSLDHDLL
SVKKQELIESTSRKLOVREARESLLEDVQANTVLGAEVAIKVGCKPSEFDFKRMF
IGDLKDVNLLLSGLSLARVENALNLDGASPGDRCSLLEKQKVLIQCHEDAELK
ENLDRRRIIVFDILANLYLSRESLADYEHFVYMKSAALLIEQRELEDKLHLEEQKCLL
DSUQPERGK"
complement(17976. .18074)
/rpt_family="HY3"
18162. .18473
/rpt_family="AluJb"
19636. .19717
/rpt_family="MER21A"
complement(19798. .19937)
/rpt_family="L1MB3"
complement(20181. .20327)
/rpt_family="FRAM"
complement(22367. .22401)
/rpt_family="At-rich"
22406. .22461
/rpt_family=" (GGAA)n"
complement(22462. .22627)
/rpt_family="AluJo"
complement(22638. .22933)
/rpt_family="AluX"
complement(26881. .26957)
/rpt_family="MERSA"
complement(27726. .27762)
/rpt_family="POLY A"
complement(27876. .27977)
/rpt_family="Aluuo/FRAM"
```

Query Match 90.6%; Score 15.4; DB 5; Length 196542;  
Best Local Similarity 94.1%; Pred. No. 2.3e+03;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGCCGAG 17  
|||||  
Db 17324 GCCCATGCTTGCCGAG 17308

RESULT 85  
AC005054/c  
LOCUS AC005054 197456 bp DNA linear HTG 17-AUG-2000  
DEFINITION Homo sapiens chromosome UNK clone CTB-45HS, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, 18 unordered pieces.  
ACCESSION AC005054  
VERSION AC005054.2 GI:9838352  
KEYWORDS HTG; HTGS\_PHASE1.

```
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 197456)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
REFERENCE 2 (bases 1 to 197456)
Waterston,R.H.
Direct Submission
Submitted (12-JUN-1998) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Aug 17, 2000 this sequence version replaced gi:3212937.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site:http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H RG045H05
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-primer ET; 71% of reads
Chemistry: Dye-terminator Big Dye; 29% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: bases at least Q40
Consensus quality: bases at least Q30
Consensus quality: bases at least Q20
Insert size: 181000; agarose-fp
Insert size: 195258; sum-of-contigs
Quality coverage: in Q20 bases; agarose-fp
Quality coverage: in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 5084: contig of 5084 bp in length
* 5085 5184: gap of unknown length
* 5185 14399: contig of 9215 bp in length
* 14400 14499: gap of unknown length
* 14500 22837: contig of 8337 bp in length
* 22837 22936: gap of unknown length
* 22936 33373: contig of 10437 bp in length
* 33373 33473: gap of unknown length
* 33473 47419: contig of 13946 bp in length
* 47419 47519: gap of unknown length
* 47520 66575: contig of 19056 bp in length
* 66575 81990: contig of 15315 bp in length
* 81991 82091: gap of unknown length
* 82091 105847: contig of 23757 bp in length
* 105848 123271: contig of 23324 bp in length
* 123271 129371: gap of unknown length
* 129371 164242: contig of 34871 bp in length
* 164243 164342: gap of unknown length
* 164343 164840: contig of 498 bp in length
* 164841 164940: gap of unknown length
* 164941 167875: contig of 2935 bp in length
* 167876 167975: gap of unknown length
* 167976 172505: contig of 4530 bp in length
* 172506 172605: gap of unknown length
* 172606 176804: contig of 4199 bp in length
* 176805 176904: gap of unknown length
```

\* 176905 180960: contig of 4056 bp in length  
 \* 180961 181060: gap of unknown length  
 \* 181061 185808: contig of 4748 bp in length  
 \* 185809 185908: gap of unknown length  
 \* 185909 190832: contig of 4924 bp in length  
 \* 190833 190932: gap of unknown length  
 \* 190933 197456: contig of 6524 bp in length.

## FEATURES

Location/Qualifiers  
 1..197456  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="UNK"  
 /clone="CTB-45H5"  
 1..5084  
 /note="assembly\_name:Contig100"  
 5085..5184  
 /estimated\_length=unknown  
 5185..14339  
 /note="assembly\_name:Contig101"  
 14400..14499  
 /estimated\_length=unknown  
 14500..22836  
 /note="assembly\_name:Contig103"  
 22837..22936  
 /estimated\_length=unknown  
 22937..33373  
 /note="assembly\_name:Contig104"  
 33374..33473  
 /estimated\_length=unknown  
 33474..47419  
 /note="assembly\_name:Contig105"  
 47420..47519  
 /estimated\_length=unknown  
 47520..66575  
 /note="assembly\_name:Contig106"  
 66576..66675  
 /estimated\_length=unknown  
 66676..81590  
 /note="assembly\_name:Contig107"  
 81591..82090  
 /estimated\_length=unknown  
 82091..105847  
 /note="assembly\_name:Contig108"  
 105848..105947  
 /estimated\_length=unknown  
 105948..129271  
 /note="assembly\_name:Contig109"  
 129272..129371  
 /estimated\_length=unknown  
 129372..164242  
 /note="assembly\_name:Contig110"  
 164243..164342  
 /estimated\_length=unknown  
 164343..164840  
 /note="assembly\_name:Contig18"  
 164841..164940  
 /estimated\_length=unknown  
 164941..167875  
 /note="assembly\_name:Contig91"  
 167876..167975  
 /estimated\_length=unknown  
 167976..172505  
 /note="assembly\_name:Contig94"  
 172506..172605  
 /estimated\_length=unknown  
 172606..176804  
 /note="assembly\_name:Contig95"  
 176805..176904  
 /estimated\_length=unknown  
 176905..180960  
 /note="assembly\_name:Contig96"  
 180961..181060

misc\_feature 181061..185808  
 /note="assembly\_name:Contig97"  
 gap 185809..185908  
 /estimated\_length=unknown  
 misc\_feature 185909..190832  
 /note="assembly\_name:Contig98"  
 gap 190833..190932  
 /estimated\_length=unknown  
 misc\_feature 190933..197456  
 /note="assembly\_name:Contig99"  
 ORIGIN  
 Query Match 90.6%; Score 15.4; DB 12; Length 197456;  
 Match Local Similarity 94.1%; Pred. No. 2.3e+03;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GCCCAGCGTTGGCCGAG 17  
 |||||  
 Db 131397 GCCCAGCGTTGGCCGAG 131381  
 |||||  
 RESULT 86  
 AC120873/c  
 LOCUS 198375 bp DNA linear ROD 15-JUN-2004  
 DEFINITION Mus musculus chromosome 12, clone RP23-283P14, complete sequence.  
 ACCESSION AC120873  
 VERSION AC120873.7 GI:48717694  
 KEYWORDS HTG.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muroidea; Muridae; Murinae; Mus.  
 Birren,B., Nusbaum,C. and Lander,E.  
 1 (bases 1 to 198375)  
 Mus musculus chromosome 12, clone RP23-283P14  
 Unpublished  
 2 (bases 1 to 198375)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
 Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,  
 Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,  
 Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,  
 Cook,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
 Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,  
 Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
 Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
 Kamat,A., Karatas,A., Kellis,C., LaRocque,K., Lamazares,R.,  
 Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,  
 MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,  
 McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Meneus,L.,  
 Minova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,  
 Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,  
 Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,  
 Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,  
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,  
 Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
 Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,  
 Topham,K., Travers,M., Travis,N., Trigglio,J., Vassiliev,H.,  
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (13-MAY-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 198375)  
 Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,  
 Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,  
 Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,  
 Collymore,A., Cook,A., Cooke,P., Corum,B., DeArelano,K.,  
 Diaz,J.S., Dodge,S., Doolley,K., Dorris,L., Erickson,J., Faro,S.,  
 Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,  
 Graham,L., Grand-Pierre,N., Hagos,B., Hagopian,D., Hagos,B.,  
 Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,



```

repeat_region complement(30641..30804)
repeat_region /rpt_family="Lx2"
repeat_region complement(30847..30950)
repeat_region /rpt_family="Lx2"
repeat_region complement(31085..32311)

Query Match
Best Local Similarity 90.6%; Score 15.4; DB 6; Length 198375;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGGCCGAG 17
|||
Db 23971 GCACACGCTTGGCCGAG 23955

RESULT 87
AC009061 AC009061 200618 bp DNA linear PRI 26-JUN-2001
DEFINITION Homo sapiens chromosome 16 clone RP11-297D21, complete sequence.
ACCESSION AC009061
VERSION AC009061.10 GI:14550301
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 200618)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
Unpublished
REFERENCE 2 (bases 1 to 200618)
DOE Joint Genome Institute.
Direct Submission
Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 200618)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
Submitted (28-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On June 26, 2001 this sequence version replaced gi:9954595.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.7% of Sequence;
Estimated Total Number of Errors is 0.4.
STS Content:
SHGC-13828 G14855
SHGC-33002 G27705
FEATURES
source Location/Qualifiers
1..200618
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-297D21"

ORIGIN
Query Match 90.6%; Score 15.4; DB 5; Length 200618;
Best Local Similarity 94.1%; Pred. No. 2.3e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGGCCGAG 17
|||
Db 94450 GCCACGCTTGGCCGAG 94466

RESULT 88
AC119828 AC119828 205915 bp DNA linear ROD 05-NOV-2003
LOCUS AC119828
DEFINITION Mus musculus chromosome 10, clone RP23-186A24, complete sequence.

```

---

```

AC119828
AC119828.9 GI:38175966
HTG.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.
1 (bases 1 to 205915)
Birren,B., Nusbaum,C. and Lander,E.
Mus musculus chromosome 10, clone RP23-186A24
Unpublished
2 (bases 1 to 205915)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,
Boukhgalter,B., Brown,A., Camarata,J., Campioano,A., Chang,J.,
Chazaro,B., Choepel,Y., Collangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Larocque,K., Lamazares,R.,
Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McGowan,P., McKernan,K., Meldrim,J., Meneus,L.,
Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (02-MAY-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 205915)
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collymore,A., Cook,A., Cooke,P., Corum,B., DeArelano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Lui,X., Mabbitt,R., MacLean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
McDonald,J., Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (11-OCT-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 205915)
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collymore,A., Cook,A., Cooke,P., Corum,B., DeArelano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Lui,X., Mabbitt,R., MacLean,C.,

```

Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,  
 Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,  
 Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,  
 O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,  
 Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,  
 Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,  
 Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,  
 Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,  
 Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,  
 Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (05-NOV-2003) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Nov 5, 2003 this sequence version replaced gi:37620325.  
 All repeats were identified using RepeatMasker:  
 Snit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: L19587

Center clone name: 186\_A\_24

----- Location/Qualifiers

1. 205915  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /chromosome="10"  
 /map="10"

/clone\_lib="RPCI-23 Female Mouse BAC"

1. 2187

/note="wgs\_end\_extension"

clone\_end:SP6

/note="complement(34..38)

/note="<30 qual SNGL region"

complement(92..96)

/note="<30 qual SNGL region"

complement(906..1320)

/rpt\_family="MTB"

1326..1424

/rpt\_family="B4A"

1804..1850

/rpt\_family="(CCA)n"

1917..1946

/rpt\_family="(CAGA)n"

2188..2193

/note="clone boundary"

clone\_end:SP6

site:EcoRI

2632..2723

/rpt\_family="L2"

complement(3241..3331)

/rpt\_family="MIR"

4268..4292

/rpt\_family="(CA)n"

4510..4688

/rpt\_family="L1ME"

5052..5421

/rpt\_family="Lx8"

5435..5457

/rpt\_family="(TTTG)n"

5467..5810

/rpt\_family="Lx8"

5884..5959

/rpt\_family="RMER15"

6254..6276

/rpt\_family="AT\_rich"

6278..6427

repeat\_region  
 /rpt\_family="B4A"  
 6599..6718  
 /rpt\_family="MER5A"  
 7692..7968  
 /rpt\_family="B4"  
 10091..10172  
 /rpt\_family="RLTR14"  
 11372..11508  
 /rpt\_family="B1\_MM"  
 11509..11535  
 /rpt\_family="B4"  
 12285..12405  
 /rpt\_family="L1MA4A"  
 12591..12643  
 /rpt\_family="CT-rich"  
 12789..12901  
 /rpt\_family="L1MA4A"  
 12941..12966  
 /rpt\_family="(TTA)n"  
 12981..13045  
 /rpt\_family="(TG)n"  
 complement(13048..13169)  
 /rpt\_family="RSINE1"  
 13170..13311  
 /rpt\_family="L1MA4A"  
 13523..13549  
 /rpt\_family="AT\_rich"  
 complement(14629..14807)  
 /rpt\_family="URR1B"  
 14969..15011  
 /rpt\_family="(TG)n"  
 complement(15416..15506)  
 /rpt\_family="MIR"  
 15658..15680  
 /rpt\_family="(CAGA)n"  
 15940..16034  
 /rpt\_family="Lx2"  
 16564..16926  
 /rpt\_family="ORR1B2"  
 complement(17275..17583)  
 /rpt\_family="ORR1A3"  
 complement(18440..18617)  
 /rpt\_family="B4A"  
 complement(18632..18834)  
 /rpt\_family="URR1A"  
 19286..19807  
 /rpt\_family="L1MB8"  
 19808..20133  
 /rpt\_family="ORR1A2"  
 20134..20527  
 /rpt\_family="L1MB8"  
 20884..20913  
 repeat\_region

Query Match 90.6%; Score 15.4; DB 6; Length 205915;

Best Local Similarity 94.1%; Pred. No. 2.4e+03;

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCCGAG 17

|||||

Db 199705 GCCCAGCTTGGCCGAG 199721

RESULT 89

AC021091/c

LOCUS

DEFINITION

AC021091

SEQUENCE, 8 ordered pieces.

AC021091

VERSION

AC021091.5

KEYWORDS

HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_ACTIVEFIN.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AC021091

Homo sapiens chromosome 16 clone RP11-421N4, WORKING DRAFT

209262 bp DNA linear HTG 25-APR-2001



Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
1 (bases 1 to 209262)  
DOE Joint Genome Institute.  
Sequencing of Human Chromosome 16  
Unpublished  
2 (bases 1 to 209262)  
DOE Joint Genome Institute.  
Direct Submission  
Submitted (14-JAN-2000) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
On Apr 25, 2001 this sequence version replaced gi:9954658.  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov

-----  
Project Information  
Center Project Name: 568345  
Center clone name: RPCI-11\_421N4  
-----

Summary Statistics  
Consensus quality: 201548 bases at least Q40  
Consensus quality: 207017 bases at least Q30  
Consensus quality: 207950 bases at least Q20  
Estimated insert size: 238990; agarose-fp estimation  
Estimated insert size: 208912; sum-of-contigs estimation  
Quality coverage: 6.28 in Q20 bases; agarose-fp estimation  
Quality coverage: 7.19 in Q20 bases; sum-of-contigs estimation.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 8 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.

\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

\* 1 5137: contig of 5137 bp in length  
\* 5138 5237: gap of unknown length  
\* 5238 47243: contig of 42006 bp in length  
\* 47244 47343: gap of unknown length  
\* 47344 68342: contig of 20999 bp in length  
\* 68343 68442: gap of unknown length  
\* 68443 151499: contig of 83057 bp in length  
\* 151500 151599: gap of unknown length  
\* 151600 161329: contig of 9730 bp in length  
\* 161330 163824: contig of 2395 bp in length  
\* 161430 163825: gap of unknown length  
\* 163825 202407: contig of 38483 bp in length  
\* 163925 202407: contig of 38483 bp in length  
\* 202408 202507: gap of unknown length  
\* 202508 209262: contig of 6755 bp in length.

FEATURES  
source

1. 209262  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/clone="RP11-421N4"  
/clone\_lib="RPCI human BAC library 11"

5138..5237  
/estimated\_length=unknown  
47244..47343  
/estimated\_length=unknown  
68343..68442  
/estimated\_length=unknown  
151500..151599  
/estimated\_length=unknown  
161330..161429  
/estimated\_length=unknown  
163825..163924  
/estimated\_length=unknown

gap  
ORIGIN

Query Match 90.6%; Score 15.4; DB 12; Length 209262;  
Best Local Similarity 94.1%; Pred. No. 2.4e+03;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCCAGCGCTTGGCGAG 17  
|||||

Db 80342 GCCCAGCGCTTGGCGAG 80326

RESULT 90

AC091522

LOCUS

DEFINITION

AC091522

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

On Nov 27, 2002 this sequence version replaced gi:24961369.

Center: Department Of Chemistry And Biochemistry

The University Of Oklahoma

Center code:UOKNOR

-----

FEATURES

source

1. 214798

/organism="Mus musculus"

/mol\_type="genomic DNA"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/chromosome="10"

/clone="rp23-199g2"

/clone\_lib="RPCI - 23 Female (C57BL/6J) Mouse BAC Library"

ORIGIN

Query Match 90.6%; Score 15.4; DB 6; Length 214798;

Best Local Similarity 94.1%; Pred. No. 2.4e+03;

```

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCACGCTTGGCGAG 17
|||||
Db 50950 GCCACGCTTGGCCAG 50966

RESULT 91
AC134138
LOCUS
DEFINITION
Rattus norvegicus clone CH230-257F9, *** SEQUENCING IN PROGRESS
AC134138 216540 bp DNA linear HTG 15-NOV-2002
***, 3 unordered pieces.
AC134138 GI:25007343
VERSION
HTG; HTGS PHASE1; HTGS_ENRICHED.
KEYWORDS
Rattus norvegicus (Norway rat)
SOURCE
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Rattus.
1 (bases 1 to 216540)
Muzny,D,Marle., Metzker,M, Lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Ayalabechechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroil,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Evans,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamill,C., Hamilton,K., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hoques,M.,
Hollins,B., Howells,S., Hui,Y., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Loulsegod,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidas,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokenleh,O., Okwuonu,G., Olarunpusagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.,
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Rilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajd,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,O., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,X., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 216540)

AUTHORS
TITLE
JOURNAL
REFERENCE

Rat Genome Sequencing Consortium.
Direct Submission
Submitted (23-SEP-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 216540)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 15, 2002 this sequence version replaced gi:23269344.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: KBWU
Center clone name: CH230-257F9
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 196236 bases at least Q40
Consensus quality: 199589 bases at least Q30
Consensus quality: 201566 bases at least Q20
Estimated insert size: 203201; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
consists of 3 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
1 212460: contig of 212460 bp in length
212461 212560: gap of unknown length
212561 213819: contig of 1259 bp in length
213820 213919: gap of unknown length
213920 216540: contig of 2621 bp in length.
FEATURES
source
Location/Qualifiers
1..216540
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-257F9"
misc_feature
1..1135
/note="wgs_end_extension
clone_end:T7"
5118..5953
/note="clone_boundary
clone_end:T7"
site:
end sequence:BZ141658"
130387..132038
/note="wgs contig"
208312..209023
/note="clone_boundary"
TITLE
JOURNAL
REFERENCE

```

```
clone_end:Sp6
site:
end_sequence:B2141659"
CR974487 .212560
/estimated_length=unknown
213820 .213919
/estimated_length=unknown

ORIGIN

Query Match          90.6%; Score 15.4; DB 12; Length 216540;
Best Local Similarity 94.1%; Pred. No. 2.4e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGCCGAG 17
|||
Db 167143 GCCCAGCTTGCCGAG 167159

RESULT 92
CR974487
LOCUS
DEFINITION Mouse DNA sequence from clone RP23-136M7 on chromosome 12, complete
sequence.
ACCESSION CR974487
VERSION CR974487.10 GI:82465399
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 218303)
Hunter,G.
Direct Submission
Submitted (15-NOV-2005) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
Clone requests: clonerequest@sanger.ac.uk
On Nov 15, 2005 this sequence version replaced gi:78271849.
-----Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: vegas@sanger.ac.uk
-----
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one subclone; and the assembly was confirmed by restriction digest,
except on the rare occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
Sequence from the Mouse Genome Sequencing Consortium whole genome
shotgun may have been used to confirm this sequence. Sequence data
from the whole genome shotgun alone has only been used where it has
a phred quality of at least 30.
RP23-136M7 is from the RPCI-23 Mouse BAC Library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6.
Location/Qualifiers
1. 218303
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="12"
/clone="RP23-136M7"
/clone_lib="RPCI-23"

ORIGIN

Query Match          90.6%; Score 15.4; DB 6; Length 218303;
Best Local Similarity 94.1%; Pred. No. 2.4e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGCCGAG 17
|||
Db 218080 GCACAGCTTGCCGAG 218096

RESULT 93
AC150072
LOCUS
DEFINITION Gallus gallus clone CH261-91K4, WORKING DRAFT SEQUENCE, 10 ordered
pieces.
ACCESSION AC150072
VERSION AC150072.1 GI:49533667
KEYWORDS HTG; HTGS PHASE2; HTGS DRAFT; HTGS_ACTIVEFIN.
SOURCE Gallus gallus (chicken)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE
1 (bases 1 to 230993)
DOE Joint Genome Institute.
Unpublished
TITLE
JOURNAL
REFERENCE
2 (bases 1 to 230993)
DOE Joint Genome Institute.
Direct Submission
Submitted (01-JUL-2004) Production Genomics Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA
94598-1638, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 2965963
Center clone name: JF2-91K4
-----
Summary Statistics
Consensus quality: 227292 bases at least Q40
Consensus quality: 228380 bases at least Q30
Consensus quality: 229249 bases at least Q20
Estimated insert size: 218000; agarose-fp estimation
Estimated insert size: 230093; sum-of-contigs estimation
Quality coverage: 11.13 in Q20 bases; agarose-fp estimation
Quality coverage: 10.54 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 49795: contig of 49795 bp in length
* 49796 49895: gap of unknown length
* 49896 104846: contig of 54951 bp in length
* 104847 120130: contig of 15184 bp in length
* 120131 129521: contig of 9291 bp in length
* 129522 129621: gap of unknown length
* 129622 154193: contig of 24572 bp in length
* 154194 154293: gap of unknown length
* 154294 157923: contig of 3630 bp in length
* 157924 158023: gap of unknown length
* 158024 160903: contig of 2880 bp in length
* 160904 161003: gap of unknown length
```

\* 161004 173646: contig of 12643 bp in length  
\* 173647 173746: gap of unknown length  
\* 173747 190902: contig of 17156 bp in length  
\* 190903 191002: gap of unknown length  
\* 191003 230993: contig of 39991 bp in length.

FEATURES  
source

Location/Qualifiers  
1..230993  
/organism="Gallus gallus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9031"  
/clone="CH261-91K4"  
/clone\_lib="CHORI-261 Chicken BAC Library"  
49796..49895  
/estimated\_length=unknown  
104847..104946  
/estimated\_length=unknown  
120131..120230  
/estimated\_length=unknown  
129522..129621  
/estimated\_length=unknown  
154194..154293  
/estimated\_length=unknown  
157924..158023  
/estimated\_length=unknown  
160904..161003  
/estimated\_length=unknown  
173647..173746  
/estimated\_length=unknown  
190903..191002  
/estimated\_length=unknown

## ORIGIN

Query Match 90.6%; Score 15.4; DB 12; Length 230993;  
Best Local Similarity 94.1%; Pred. No. 2.4e+03;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCCTGGCCGAG 17  
|||||

Db 135912 GCCCAGCCTGGCCGAG 135896

## RESULT 94

AC150066 233345 bp DNA linear HTG 01-JUL-2004  
LOCUS  
DEFINITION Gallus gallus clone CH261-67M3, WORKING DRAFT SEQUENCE, 7 ordered  
pieces.

## ACCESSION

AC150066 1 GI:49533661  
VERSION  
HTG; HTGS PHASE2; HTGS\_DRAFT; HTGS\_ACTIVEFIN.

## KEYWORDS

SOURCE  
Gallus gallus (Chicken)  
Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus

## REFERENCE

1 (bases 1 to 233345)  
DOE Joint Genome Institute.

## AUTHORS

Unpublished

## TITLE

Unpublished

## REFERENCE

2 (bases 1 to 233345)  
DOE Joint Genome Institute.  
Direct Submission  
Submitted (01-JUL-2004) Production Genomics Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA  
94598-1698, USA

## COMMENT

-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: <http://www.jgi.doe.gov>  
-----

Project Information  
Center Project Name: 2956794

Center clone name: JF2-67M3

-----

Summary Statistics  
Consensus quality: 230345 bases at least Q40  
Consensus quality: 231164 bases at least Q30  
Consensus quality: 231877 bases at least Q20  
Estimated insert size: 248000; agarose-fp estimation  
Estimated coverage: 12.16 in Q20 bases; sum-of-contigs estimation  
Quality coverage: 12.96 in Q20 bases; agarose-fp estimation  
NOTE: This is a 'working draft' sequence. It currently  
consists of 7 contigs. Gaps between the contigs  
are represented as runs of N. The order of the pieces  
is believed to be correct as given, however the sizes  
of the gaps between them are based on estimates that have  
been provided by the submitter.

\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

1 159426: contig of 159426 bp in length  
\* 159427 159526: gap of unknown length  
\* 159527 162471: contig of 2945 bp in length  
\* 162472 162571: gap of unknown length  
\* 162572 166152: contig of 3581 bp in length  
\* 166153 166252: gap of unknown length  
\* 166253 183843: contig of 17591 bp in length  
\* 183844 183943: gap of unknown length  
\* 183944 182052: contig of 8109 bp in length  
\* 192053 192152: gap of unknown length  
\* 192153 217751: contig of 25599 bp in length  
\* 217752 217851: gap of unknown length  
\* 217852 233345: contig of 15494 bp in length.

FEATURES  
source

1..233345  
/organism="Gallus gallus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9031"  
/clone="CH261-67M3"  
/clone\_lib="CHORI-261 Chicken BAC Library"  
159427..159526  
/estimated\_length=unknown  
162472..162571  
/estimated\_length=unknown  
166153..166252  
/estimated\_length=unknown  
183844..183943  
/estimated\_length=unknown  
192053..192152  
/estimated\_length=unknown  
217752..217851  
/estimated\_length=unknown

## ORIGIN

Query Match 90.6%; Score 15.4; DB 12; Length 233345;  
Best Local Similarity 94.1%; Pred. No. 2.4e+03;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCCTGGCCGAG 17  
|||||

Db 185142 GCCCAGCCTGGCCGAG 185158

## RESULT 95

## AC130741

## LOCUS

DEFINITION AC130741 234079 bp DNA linear HTG 10-MAY-2003  
Rattus norvegicus clone CH230-84A14, \*\*\* SEQUENCING IN PROGRESS  
\*\*\*, 4 unordered pieces.

## ACCESSION

AC130741 4 GI:30522016

## VERSION

HTG; HTGS PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.

## KEYWORDS

SOURCE Rattus norvegicus (Norway rat)

## ORGANISM

Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muroidea; Muridae; Rattus.

REFERENCE  
AUTHORS

1 (bases 1 to 234079)  
Muzny, D. Marie, Metzker, M. Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsebrook, S., Amin, A., Anguiano, D., Anyalebechi, V., Ayodeji, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biwalo, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Frazer, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregorjis, E., Geis, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, M., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuwa, L., Loulisedge, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Mahidatne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokemele, O., Okwuonu, G., Olarpunsgoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quilroz, J., Renlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Reilly, E., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C. D., Smajda, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Vallar, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

TITLE  
JOURNAL

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

2 (bases 1 to 234079)  
Rat Genome Sequencing Consortium.  
Submitted (14-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

3 (bases 1 to 234079)  
Rat Genome Sequencing Consortium.  
Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT

On May 10, 2003 this sequence version replaced gi:25009391. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence

contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GGBE  
Center clone name: CH230-84A14  
----- Summary Statistics  
Assembly program: Atlas 3.0;  
Consensus quality: 208845 bases at least Q40  
Consensus quality: 213470 bases at least Q30  
Consensus quality: 216683 bases at least Q20  
Estimated insert size: 217164; sum-of-contigs estimation  
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation  
-----

\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 4 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence.  
\* as soon as it is available and the accession number will  
\* be preserved.

1 90552: contig of 90552 bp in length  
\* 90553 90652: gap of unknown length  
\* 90653 225351: contig of 134699 bp in length  
\* 225352 225451: gap of unknown length  
\* 225452 230188: contig of 4737 bp in length  
\* 230189 230288: gap of unknown length  
\* 230289 234079: contig of 3791 bp in length.

FEATURES  
source

1. 234079  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clones="CH230-84A14"

misc\_feature

1. 1516  
/note="wgs contig"

gap

90553..90652  
/estimated\_length=unknown

misc\_feature

148829..150032  
/note="wgs contig"

misc\_feature

219786..220971  
/note="wgs contig"

gap

225352..225451  
/estimated\_length=unknown

gap

230189..230288  
/estimated\_length=unknown

ORIGIN

Query Match 90.6%; Score 15.4; DB 12; Length 234079;  
Best Local Similarity 94.1%; Pred. No. 2.4e+03;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCACGCTTGGCCGAG 17  
|||||

Db 110681 GCCCATGCTTGGCCGAG 110697  
|||||

RESULT 96

AC165587

LOCUS

AC165587

DEFINITION

Bos taurus clone CH240-169117, \*\*\* SEQUENCING IN PROGRESS \*\*\*

ACCESSION

AC165587

VERSION

AC165587.2 GI:85664371

236105 bp DNA linear HTG 23-JAN-2006

HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.  
 SOURCE Bos taurus (cattle)  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.  
 1 (bases 1 to 236105)  
 REFERENCE Muzny, D., Adams, C., Agbai II, O., Allen, C., Alsbrooks, S., Archer, P., Arredondo, H., Bandaranaike, D., Bangura, L., Beltran, B., Beltran, R., Beraducci, A., Biswal, K., Blyth, P., Bonham, K., Buhay, C., Burch, P., Cadoree, I., Canada, A., Cardenas, V., Carter, K., Cavazos, I., Chacko, J., Chahrour, M., Chavez, D., Chen, A., Chen, G., Chen, R., Cheng, M.-T., Chu, J., Clerc, K., Cockrell, R., Coyle, M., Cree, A., Curry, S., Dai, W., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinth, H., Donlin, J., McCauley, S., Dugan-Rocha, S., Dunn, A., Durbin, K., Diuadi, D., Egan, A., Escotto, M., Espinosa, V., Eugene, C., Fa, M., Fernandez, S., Fernando, P., Flagg, N., Forbes, L., Foster, P., Fowler, G., Fu, Q., Fuh, E., Garcia, A., Garcia, R., Garner, T., Gaskin, C., Gench, S., Ghose, S., Gill, R., Gonzalez, D., Gonzalez-Garay, M., Guevara, W., Holder, M., Haaland, W., Hall, B., Hamid, H., Hamilton, K., Harbes, B., Harris, R., Havlak, P., Hawes, A., Hawkins, E., Hayes, S., Hemphill, L., Hernandez, J., Hines, S., Hitchens, M., Hodgson, A., Hognes, M., Hollins, B., Howell, L. T., Hulyk, S., Hume, J., Imo, K., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Kalatus, K., Kelly, S., Keys, T., Khan, Z., King, L., Kovar, C., Kowis, A., Kowis, C., Lara, F., Leal, S., Lee, K., Lee, S., LeGall, F. I., Lemon, S., Lewis, L., Li, B., Li, Y., Li, Z., Linnell, M., Liu, W., Liu, Y.-S., Liu, Y., Liyanage, D., London, P., Lopez, J., Lorensueta, L., Lozada, R., Luk, T., Madu, R., Maheshwari, M., Mahoney, C., Malloy, K., Mansouri, D., Martinez, E., McClelland, H., McPherson, J., Mercadado, C., Mettaker, M., Milosavljevic, A., Minja, E., Morgan, M., Morris, S., Munidasa, M., Murray, D., Nazareth, L., Ngo, D., Nguyen, N., Norwig-Eastough, E., Nott, A., Nwaekemele, O., Obregon, M., Ochi-Okorie, C., Odeh, E., Okwuonu, G., Okwuonu, K., Parker, D., Pasternak, S., Patel, B., Patel, V., Paul, H., Perez, A., Perez, L., Petrosino, J., Pham, T., Primus, E., Pu, L.-L., Puazo, M., Qin, X., Quinn, A., Quiroz, J., Rabata, D., Rachlin, E., Reigh, R., Ren, Y., Reuter, M., Richards, S., Rives, C., Rodriguez, F., Rojas, A., Ruiz, S. J., Sana, M., Sanders, W., Santibanez, J., Santos, R., Savery, G., Scherer, S., Shen, H., Shen, Y., Sisson, I., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Svatek, A., Taylor, E., Taylor, T., Thomas, N., Thorn, R., Thornton, R., Trejos, Z., Usmani, K., Vargo, C., Verdusco, D., Villasana, D., Virk, D., Volkov, A., Waldron, L., Walker, B., Wang, Q., Wang, S., Warren, J., Wei, X., Wheeler, D., Williams, G., Williams, K., Worley, K., Wright, R., Wu, J., Yakub, S., Yan, K., Yuan, Y., Yu, F., Zhang, J., Zhang, L., Zhang, Z., Zhou, J., Weinstock, G. and Gibbs, R. A.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 236105)  
 REFERENCE Worley, K. C.  
 TITLE Direct Submission  
 JOURNAL Submitted (17-JUL-2005) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 236105)  
 REFERENCE .  
 AUTHORS Bovine Genome Sequencing Consortium  
 TITLE Direct Submission  
 JOURNAL Submitted (23-JAN-2006) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 COMMENT On Jan 23, 2006 this sequence version replaced gi:70912605.  
 The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole

genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: FJWO  
 Center clone name: CH240-169117  
 ----- Summary Statistics  
 Assembly program: Atlas 3.0;  
 Consensus quality: 211571 bases at least Q40  
 Consensus quality: 213978 bases at least Q30  
 Consensus quality: 216024 bases at least Q20  
 Estimated insert size: 216034; sum-of-contigs estimation  
 Quality coverage: 5x in Q20 bases; sum-of-contigs estimation  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 30 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1  
 \* 5219: contig of 5219 bp in length  
 \* 5263: gap of 50 bp  
 \* 5270: contig of 3405 bp in length  
 \* 8675: gap of unknown length  
 \* 10213: contig of 1439 bp in length  
 \* 10263: gap of 50 bp  
 \* 10264: contig of 1714 bp in length  
 \* 17437: gap of 50 bp  
 \* 17438: contig of 53453 bp in length  
 \* 17488: gap of 50 bp  
 \* 70991: contig of 5925 bp in length  
 \* 70991: contig of 5925 bp in length  
 \* 76915: gap of 50 bp  
 \* 76916: contig of 1446 bp in length  
 \* 78411: contig of 1446 bp in length  
 \* 78412: gap of unknown length  
 \* 80033: contig of 1521 bp in length  
 \* 81311: gap of 1279 bp  
 \* 93774: contig of 18463 bp in length  
 \* 99775: gap of 306 bp  
 \* 100081: contig of 1 bp in length  
 \* 100082: gap of 77 bp  
 \* 100158: contig of 1 bp in length  
 \* 100159: gap of 50 bp  
 \* 100209: contig of 4133 bp in length  
 \* 104342: gap of 50 bp  
 \* 104393: contig of 5277 bp in length  
 \* 104392: gap of 50 bp  
 \* 104393: contig of 5277 bp in length  
 \* 109670: gap of 50 bp  
 \* 109720: contig of 4990 bp in length  
 \* 114710: gap of 50 bp  
 \* 114759: contig of 3073 bp in length  
 \* 117832: contig of 2359 bp  
 \* 120191: contig of 9937 bp in length  
 \* 130128: gap of 50 bp  
 \* 130178: contig of 2197 bp in length  
 \* 130179: gap of 7103 bp  
 \* 132376: contig of 9385 bp in length  
 \* 132376: gap of 206 bp  
 \* 148864: contig of 43016 bp in length  
 \* 149070: contig of 43016 bp in length  
 \* 192085: gap of unknown length  
 \* 192185: contig of 25591 bp in length  
 \* 192776: gap of 635 bp  
 \* 218411: contig of 1344 bp in length  
 \* 218412: gap of 50 bp  
 \* 219805: contig of 1037 bp in length  
 \* 219806: gap of unknown length  
 \* 220942: gap of unknown length



shotgun sequence only contigs will be indicated in the feature table.

```

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GCNS
Center clone name: CH230-703
----- Summary Statistics
Assembly program: Atlas;
Consensus quality: 227364 bases at least Q40
Consensus quality: 223046 bases at least Q30
Consensus quality: 230029 bases at least Q20
Estimated insert size: 235940: sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 240430: contig of 240430 bp in length.
Location/Qualifiers
1. 240430
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-703"
1. 1676
/note="wgs contig"
184225. 186084
/note="wgs contig"
237584. 240430
/note="wgs contig"

FEATURES
source
misc_feature
misc_feature
misc_feature

ORIGIN
Query Match 90.6%; Score 15.4; DB 12; Length 240430;
Best Local Similarity 94.1%; Pred. No. 2.4e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGCCGAG 17
||||| |||||
Db 238589 GCCACGCTTGCCGAG 238573

RESULT 98
AC106342
LOCUS
DEFINITION Rattus norvegicus clone CH230-154L23, WORKING DRAFT SEQUENCE, 4
unordered pieces.
AC106342
VERSION AC106342.4 GI:30581327
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Muridae; Murinae; Rattus.
1 (bases 1 to 241836)
Muzny,D,Marie., Metzker,M, Lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Ayoyi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Bunay,C., Burch,P., Burrell,K., Calderon,E.,

```

```

Cardenas,V., Carter,K., Cavazos,I., Cesar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Cartoll,L., De Anda,C., Dederich,D.,
Delgado O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebrgeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Haviak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpthy,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuewa,L., Louised,H., Lozano,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokemele,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,W., Ren,Y., Ruter,M., Richards,S., Riggs,F.,
Rivers,C., Rodkey,T., Rojars,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajd,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steinle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,I., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villalana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,R., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczky,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,X., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.

Direct Submission
Unpublished
2 (bases 1 to 241836)
Worley,K.C.
Direct Submission
Submitted (12-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 241836)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 13, 2003 this sequence version replaced gi:23111185.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM

```



Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

Center project name: GKSJ

Center clone name: CH230-154123

Assembly program: Atlas 3.0;

Consensus quality: 223364 bases at least Q40

Consensus quality: 225949 bases at least Q30

Consensus quality: 227952 bases at least Q20

Estimated insert size: 23411; sum-of-contigs estimation

Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: Estimated insert size may differ from sequence length

\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 4 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

-----

\* 1 156489: contig of 156489 bp in length

\* 156490 156589: gap of unknown length

\* 156590 238919: contig of 82230 bp in length

\* 238920 238919: gap of unknown length

\* 238920 240232: contig of 1313 bp in length

\* 240233 240332: gap of unknown length

\* 240333 241836: contig of 1504 bp in length.

#### FEATURES

source

1. 241836

/organism="Rattus norvegicus"

/db\_type="genomic DNA"

/map\_xref="taxon:10116"

/clone="CH230-154123"

1. 1042

/note="wgs contig"

156490. 156589

/estimated\_length=unknown

238920. 238919

/estimated\_length=unknown

240233. 240332

/estimated\_length=unknown

#### ORIGIN

Query Match 90.6%; Score 15.4; DB 12; Length 241836;

Best Local Similarity 94.1%; Pred. NO. 2.4e+03;

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGGCCGAG 17

Db 34630 GCCATGCTTGGCCGAG 34646

#### RESULT 99

AC111284/c

LOCUS

DEFINITION Rattus norvegicus clone CH230-5415, \*\*\* SEQUENCING IN PROGRESS \*\*\*;

2 unordered pieces.

AC111284

AC111284.3 GI:23264096

HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.

KEYWORDS Rattus norvegicus (Norway rat)

SOURCE Rattus norvegicus

ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 245117)

Murphy, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,

Allen, C., Allen, H., Alabrooks, S., Amin, A., Anguiano, D.,

Anylebeche, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,

Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,

Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,  
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,  
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, J., Chu, J.,  
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,  
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,  
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,  
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,  
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,  
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,  
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,  
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,  
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,  
Karpathy, S., Kelly, S., Khan, Z., Khan, Z., King, L., Kovar, C.,  
Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
Lorensu, L., Loulseghe, H., Lozano, R. J., Lu, X., Ma, J.,  
Maheshwari, M., Mahindartine, M., Mahmoud, M., Malloy, K., Mangum, A.,  
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,  
Mawhinney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,  
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,  
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,  
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K.,  
Nwako, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,  
Pasternak, S., Paul, H., Perez, A., Perez, L., Pflankoch, C.,  
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L. L.,  
Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,  
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.,  
Sanders, W., Savery, G., Scherer, S., Scott, G., Shattman, S., Shen, H.,  
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D.,  
Sneid, A., Sodergren, E., Song, X. Z., Sorelle, R., Sosa, J.,  
Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,  
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umami, K.,  
Valas, R., Vera, V., Villanasa, D., Waldron, L., Walker, B., Wang, J.,  
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,  
Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K.,  
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Zhao, S., Dunn, D., von  
Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,  
Weinstock, G. and Gibbs, R. A.

Direct Submission

Unpublished

2 (bases 1 to 245117)

Worley, K. C.

Direct Submission

Submitted (19-FEB-2002)

Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 245117)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (21-SEP-2002)

Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On Sep 21, 2002 this sequence version replaced gi:21735811.

The sequence in this assembly is a combination of BAC based reads

and whole genome shotgun sequencing reads assembled using Atlas

(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). As a result, the

sequence may extend beyond the ends of the clone and there may be

contigs that consist entirely of whole genome shotgun sequence

reads. Both end sequences and whole genome shotgun sequence only

contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

----- Project Information

```

Center project name: GLYH
Center clone name: CH230-54L5
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 215111 bases at least Q40
Consensus quality: 218615 bases at least Q30
Consensus quality: 221156 bases at least Q20
Estimated insert size: 236907; sum-of-contigs estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 243787: contig of 243787 bp in length
* 243788 243887: gap of unknown length
* 243888 245117: contig of 1230 bp in length.
FEATURES             Location/Qualifiers
     source            1..245117
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-54L5"
     misc_feature      1..1075
                     /note="wgs_end_extension"
                     clone_end:T7"
     misc_feature      1922..3168
                     /note="wgs_end_extension"
                     clone_end:T7"
     misc_feature      6356..8142
                     /note="wgs_end_extension"
                     clone_end:T7"
     misc_feature      8227..9080
                     /note="clone boundary"
                     clone_end:T7"
                     site:ECORI
                     end_sequence:BH350328"
     misc_feature      239598..240208
                     /note="clone boundary"
                     clone_end:Sp6
                     site:ECORI
                     end_sequence:BH350329"
     misc_feature      241850..243787
                     /note="wgs_end_extension"
                     clone_end:Sp6"
     gap              243788..243887
                     /estimated_length=unknown
ORIGIN
Query Match          90.6%; Score 15.4; DB 12; Length 245117;
Best Local Similarity 94.1%; Pred. NO. 2.4e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 GCCCAGCTTGCCGAG 17
   |||||
Db 42590 GCCCAGCTTGCCGAG 42574

RESULT 100
AC115439/c
LOCUS          AC115439          247212 bp          DNA          linear          HTG 08-OCT-2002
DEFINITION    Rattus norvegicus clone CH230-184F24, *** SEQUENCING IN PROGRESS
              *** 2 unordered pieces.
ACCESSION     AC115439
VERSION       AC115439.4  GI:23195508
KEYWORDS      HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
SOURCE        Rattus norvegicus (Norway rat)

```

## ORGANISM

Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridea; Muridae; Murinae; Rattus.

REFERENCE  
AUTHORS

1 (Bases 1 to 247212)  
 Muzny,D,Marie., Metzker,M, Lee., Abramzon,S., Adams,C., Alder,J.,  
 Allen,C., Allen,H., Alsbrooks,S., Amin,A., Angiano,D.,  
 Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,  
 Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,  
 Biwalto,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,  
 Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,  
 Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,  
 Chacko,V., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,  
 Cleveland,C., Cockrell,R., Cox,C., Coyie,M., Cree,A., D'Souza,B.,  
 Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,  
 Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,  
 Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,  
 Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,  
 Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,  
 Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,F., Garza,W.,  
 Gbregoris,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,  
 Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,  
 Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,  
 Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,  
 Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,  
 Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,  
 Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovat,C.,  
 Kowis,C., Kraft,C.L., Lebow,H., Levay,J., Lewis,L., Li,Z., Liu,J.,  
 Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,  
 Lorus,hewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,  
 Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,  
 Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,  
 Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,  
 Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,  
 Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,  
 Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,  
 Nwaokemele,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,  
 Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,  
 Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,  
 Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,  
 Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,  
 Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,  
 Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,  
 Shetty,J., Shvartsbeyn,A., Sisson,I., Sittler,C.D., Sosa,J.,  
 Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,  
 Steimle,M., Strong,R., Sutton,A., Svatek,A., Taber,P., Taylor,C.,  
 Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,  
 Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,  
 Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,  
 Williams,G., Willson,R., Wlezyk,R., Wooden,H., Worley,K.,  
 Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,  
 Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von  
 Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,  
 Weinstock,G. and Gibbs,R.A.

## Direct Submission

Unpublished

2 (Bases 1 to 247212)

Worley,K.C.

## Direct Submission

Submitted (19-MAR-2002)

of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

3 (Bases 1 to 247212)

## Rat Genome Sequencing Consortium.

Direct Submission

Submitted (08-OCT-2002)

of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

On Sep 19, 2002 this sequence version replaced gi:21737862.

The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described  
 in the feature table below represents a scaffold in the Atlas

assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

```

----- Genome Center
Center: Baylor College Of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GPWF
Center clone name: CH230-184F24
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 218155 bases at least Q40
Consensus quality: 220835 bases at least Q30
Consensus quality: 222648 bases at least Q20
Estimated insert size: 238245; sum-of-contigs estimation
Quality coverage: 3x in Q20 bases; sum-of-contigs estimation
-----

```

\* NOTE: Estimated insert size may differ from sequence length (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

* 1 115755: contig of 115755 bp in length
* 115756 115855: Gap of unknown length
* 115856 247212: contig of 131357 bp in length.

```

## FEATURES

```

source
  1..247212
    /organism="Rattus norvegicus"
    /mol_type="genomic DNA"
    /db_xref="taxon:10116"
    /clone="CH230-184F24"
  misc_feature
    1..1360
    /note="wgs contig"
  misc_feature
    114369..115755
    /note="wgs contig"
  gap
    115756..115855
    /estimated_length=unknown
  misc_feature
    200330..201623
    /note="wgs contig"

```

## ORIGIN

```

Query Match          90.6%; Score 15.4; DB 12; Length 247212;
Best Local Similarity 94.1%; Pred. No. 2.4e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Oy  1  GCCCAGCGTGGCCGAG 17
    |||||
Db  144996 GCCCAGCGTGGCCGAG 144980

```

Search completed: June 10, 2006, 18:20:43  
Job time : 518.533 secs

**THIS PAGE IS BLANK**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2006, 15:16:40 ; Search time 126.367 Seconds  
(without alignments)  
937.971 Million cell updates/sec

Title: US-10-600-816-31

Perfect score: 17

Sequence: 1 gccacgctggccgag 17

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 5244920 seqs, 3486124231 residues

Total number of hits satisfying chosen parameters: 10489840

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

Database :

N Geneseq\_8:\*

1: Geneseqn1980s:\*

2: Geneseqn1990s:\*

3: Geneseqn2000s:\*

4: Geneseqn2001as:\*

5: Geneseqn2002bs:\*

6: Geneseqn2002as:\*

7: Geneseqn2003as:\*

8: Geneseqn2003bs:\*

9: Geneseqn2003cs:\*

10: Geneseqn2003ds:\*

11: Geneseqn2004as:\*

12: Geneseqn2004bs:\*

13: Geneseqn2004cs:\*

14: Geneseqn2005s:\*

15: Geneseqn2006s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	17	100.0	17	12	ADI28528 Human GPC
2	17	100.0	40	12	ADI28527 Human GPC
3	17	100.0	497	11	Adt96418 Colon can
4	17	100.0	497	11	ADX42900 Human cdn
5	17	100.0	552	11	Adt96292 Colon can
6	17	100.0	552	11	ADX42774 Human cdn
7	17	100.0	603	3	Aaz90046 Hydrophob
8	17	100.0	620	6	ABQ58527 Human col
9	17	100.0	634	6	ABQ59698 Human col
10	17	100.0	642	10	ABT22328 Breat ca
11	17	100.0	1074	12	Ado30035 Human GPC
12	17	100.0	1212	3	Aaz61776 cDNA enco
13	17	100.0	1212	4	AAC99709 Skin cell
14	17	100.0	1212	6	ABL34861 Human cdn
15	17	100.0	1228	8	ACA04775 cDNA enco
16	17	100.0	1460	14	AEA00112 Human TAT
17	17	100.0	1460	14	AEA00632 Human TAT
18	17	100.0	1619	4	AAF58615 Human REC

19	17	100.0	1718	3	AAZ90056	Aaz90056 Hydrophob
20	17	100.0	1788	10	ADF70573	Adf70573 Orphan re
21	17	100.0	2302	6	ABT10173	Abt10173 Human bre
22	17	100.0	2302	8	ACC58386	Acc58386 Human GPC
23	17	100.0	2302	10	ADD93240	Ad93240 RAIG1 cod
24	17	100.0	2302	10	ADL24773	Adl24773 Intestina
25	17	100.0	2316	10	ABT31923	Abt31923 Human bre
26	17	100.0	2446	4	AAH14688	Aah14688 Human cdn
27	17	100.0	2456	8	ABZ42832	Abz42832 Human G p
28	17	100.0	2456	10	ACC72695	Acc72695 Human can
29	17	100.0	2456	11	ADN39301	Adn39301 Cancer/an
30	17	100.0	2456	12	ADI28525	Adi28525 Human GPC
31	17	100.0	2456	12	ADI28459	Adi28459 Human GPC
32	17	100.0	2456	13	ADR48221	Adr48221 Human ret
33	17	100.0	2456	13	ACN39388	Acn39388 Tumour-as
34	17	100.0	2456	13	ADR43992	Adr43992 Human bre
35	17	100.0	2456	13	ADU06126	Adu06126 Novel bro
36	17	100.0	2456	14	AEA00088	Aea00088 Human TAT
37	17	100.0	2456	14	AEA00608	Aea00608 Human TAT
38	17	100.0	2456	14	AED47480	Aed47480 Retinoic
39	17	100.0	2593	6	ABQ54954	Abq54954 Human ova
40	17	100.0	3371	4	AAH72766	Aah72766 Human cer
41	17	100.0	4239	11	ACN89274	Acn89274 Breast ca
42	17	100.0	6730	13	ADX97494	Adx97494 Pancreati
43	15.4	90.6	1527	11	ABD12717	Abd12717 Pseudomon
44	15.4	90.6	2306	6	ABQ70591	Abq70591 Listeria
45	15.4	90.6	2976	14	AEA48186	Aea48186 DNA of si
46	15.4	90.6	4183	2	AXH81946	Axh81946 DNA encod
47	15.4	90.6	4222	12	ADQ24610	Adq24610 Human sof
48	15.4	90.6	4232	13	ADRO7536	Adro7536 Full leng
49	15.4	90.6	9699	5	ABA21144	Aba21144 Human ner
50	15.4	90.6	17457	5	ABA21143	Aba21143 Human ner
51	15.4	90.6	17458	5	ABA21142	Aba21142 Human ner
52	15.4	90.6	63761	13	ABD33426	Abd33426 Murine ca
53	15.4	90.6	68732	13	ABD33428	Abd33428 Human can
54	15.4	90.6	95269	6	ABQ67195	Abq67195 4
55	15.4	90.6	110000	6	ABQ69245	Abq69245 13
56	15.4	90.6	110000	6	ABQ67195	Abq67195 3
57	15.4	90.6	110000	6	ABA03041	Aba03041 13
58	15.4	90.6	349980	6	ABQ81846	Abq81846 Bifidobac
59	15.4	90.6	349980	6	ABQ81847	Abq81847 Bifidobac
60	15	88.2	337	5	ABV57592	Abv57592 Human pro
61	15	88.2	3303	13	ADT19568	Adt19568 Plant cdn
62	14.4	84.7	344	4	AAS37371	Aas37371 Novel hum
63	14.4	84.7	546	4	AAH09788	Aah09788 Human cdn
64	14.4	84.7	570	5	ABV54022	Abv54022 Human pro
65	14.4	84.7	680	4	AAH08354	Aah08354 Human cdn
66	14.4	84.7	992	8	AAI55183	Aai55183 Verticill
67	14.4	84.7	1520	2	AAV63901	Aav63901 Helicobac
68	14.4	84.7	1531	8	ACD05884	Acd05884 Novel hum
69	14.4	84.7	1675	2	AAV63900	Aav63900 Helicobac
70	14.4	84.7	1889	6	ABZ35358	Abz35358 Human gen
71	14.4	84.7	3681	13	ACN40151	Acn40151 Tumour-as
72	14.4	84.7	13608	6	ABK89239	Abk89239 Human Na/
73	14.4	84.7	48169	14	AED48654	Aed48654 Streptomy
74	14.4	84.7	49312	3	AAH51594	Aah51594 Human gen
75	14.4	84.7	89378	12	ADN46845	Adn46845 20
76	14.4	84.7	89378	12	ADN46123	Adn46123 20
77	14.4	84.7	89378	12	ADN46464	Adn46464 20
78	14.4	84.7	110000	12	ADN47591	Adn47591 Thermococ
79	14.4	84.7	110000	12	ADN47209	Adn47209 Thermococ
80	14.4	84.7	110000	12	ADN47960	Adn47960 Thermococ
81	14.4	84.7	32602	13	ABD32843	Abd32843 Human can
82	14	82.4	338	6	ABN23393	Abn23393 Human ORF
83	14	82.4	458	9	ACH41474	Ach41474 Human foe
84	14	82.4	508	12	ADN12918	Adn12918 Human pro
85	14	82.4	706	12	ADN13857	Adn13857 Human pro
86	14	82.4	1234	3	AAC47788	Aac47788 Arabidops
87	14	82.4	110000	12	ADN46845	Adn46845 16
88	14	82.4	110000	12	ADN47591	Adn47591 04
89	14	82.4	110000	12	ADN46123	Adn46123 16
90	14	82.4	110000	12	ADN47209	Adn47209 04
91	14	82.4	110000	12	ADN46464	Adn46464 16















C 968	12.8	75.3	246	2	AAT26025	Aat26025	Human gen
C 969	12.8	75.3	248	7	ADS65052	Adse65052	Corn seed
C 970	12.8	75.3	248	7	ADS65016	Adse65016	Corn seed
C 971	12.8	75.3	249	6	ABN18009	Abn18009	Human ORF
C 972	12.8	75.3	251	2	AAX11441	Aax11441	Human b1a
C 973	12.8	75.3	252	10	AD136346	Adi36346	Mature hH
C 974	12.8	75.3	254	6	AD35785	Ad35785	BS203 cDN
C 975	12.8	75.3	254	14	AD359580	Ady59580	Breast ti
C 976	12.8	75.3	258	10	AD136345	Adi36345	Mature hH
C 977	12.8	75.3	259	6	ABN78224	Abn78224	Human ORF
C 978	12.8	75.3	262	6	AD35783	Ad35783	BS203 cDN
C 979	12.8	75.3	262	14	ADY59578	Ady59578	Breast ti
C 980	12.8	75.3	263	6	AD35784	Ad35784	BS203 cDN
C 981	12.8	75.3	263	14	ADY59579	Ady59579	Breast ti
C 982	12.8	75.3	264	5	ABA09609	Abao9609	Human bon
C 983	12.8	75.3	264	9	ADA47999	Ada47999	Rice gene
C 984	12.8	75.3	269	7	AD365606	Adse6506	Corn seed
C 985	12.8	75.3	270	4	AAS51110	Aas51110	Salmonell
C 986	12.8	75.3	271	14	ACS38537	Adc38537	Human can
C 987	12.8	75.3	288	4	ABL23447	Abl23447	Drosophil
C 988	12.8	75.3	312	10	ADU716344	Adi36344	Fulli-leng
C 989	12.8	75.3	320	13	ADU76397	Adu76397	DNA probe
C 990	12.8	75.3	325	2	AAH87506	Aah87506	Human sin
C 991	12.8	75.3	330	3	AAC02983	Aac02983	Human sec
C 992	12.8	75.3	330	6	ABN77832	Abn77832	Human kin
C 993	12.8	75.3	346	5	ABA09693	Abao9693	Human bon
C 994	12.8	75.3	366	4	AAI91909	Aai91909	Human pol
C 995	12.8	75.3	367	5	ABA11438	Abai11438	Human ner
C 996	12.8	75.3	373	10	ADD27316	Add27316	Human adi
C 997	12.8	75.3	379	4	AAK55424	Aak55424	Human imm
C 998	12.8	75.3	381	8	ACA24045	Aca24045	Prokaryot
C 999	12.8	75.3	384	3	AAF15257	Aaf15257	Trichoder
C 1000	12.8	75.3	384	13	ADU59298	Adu59298	Trichoder

## ALIGNMENTS

CC	function and/or regulation, E-selectin function and/or regulation, or aberrant NF-kB function and/or regulation (all claimed).
XX	
SQ	Sequence 17 BP; 2 A; 7 C; 6 G; 2 T; 0 U; 0 Other;
	Query Match 100.0%; Score 17; DB 12; Length 17;
	Best Local Similarity 100.0%; Pred. No. 59;
	Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	1 GCCACGCTTGCCGAG 17
DB	1 GCCACGCTTGCCGAG 17
RESULT 2	
ADI28527	
ID	ADI28527 standard; DNA; 40 BP.
XX	
AC	ADI28527;
XX	
DT	22-APR-2004 (first entry)
XX	
DE	Human GPCR retinoic acid induced 3 (RAI-3) peptide DNA.
XX	
KW	Retinoic acid induced 3; RAI-3; human; G-protein coupled receptor; GPCR;
KW	antiinflammatory; immunosuppressive; cytostatic; cardiant; antiallergic;
KW	broncholytic; gene therapy; PCR; primer; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO2004001060-A2.
XX	
PD	31-DEC-2003.
XX	
PF	20-JUN-2003; 2003WO-US019255.
XX	
PR	20-JUN-2002; 2002US-0390850P.
PR	29-AUG-2002; 2002US-0407006P.
XX	
PA	(BRIM ) BRISTOL-MYERS SQUIBB CO.
XX	
PI	Whitney GS, Opitack G, Garulacan L, Ramanathan CS, McKinnon M;
PI	Bennett KL, Barber LE, Cacace A, Tsuchihashi Z;
XX	
XX	WPI; 2004-090973/09.
XX	
XX	New nucleic acid molecule encoding a human G-protein coupled receptor,
PT	RAI3, useful for preventing, treating or ameliorating chronic obstructive
PT	pulmonary disease (COPD), COPD-like disorder, or the underlying symptoms
PT	

```
XX 20-JUN-2002; 2002US-0390850P.
PR 29-AUG-2002; 2002US-0407006P.
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.
PA Whitney GS, Opitack G, Garulacan L, Ramanathan CS, McKinnon M;
PI Bennett KL, Barber LE, Cacace A, Tsuchihashi Z;
XX WPI: 2004-090973/09.
DR P-PSDB; ADI28458.
XX New nucleic acid molecule encoding a human G-protein coupled receptor,
PT RAI3, useful for preventing, treating or ameliorating chronic obstructive
PT pulmonary disease (COPD), COPD-like disorder, or the underlying symptoms
PT of COPD.
PS Disclosure: SEQ ID NO 30; 301pp; English.
XX The present sequence is that of DNA encoding a peptide corresponding to
CC amino acids 340-353 of a human G-protein coupled receptor, denoted
CC retinoic acid induced 3 (RAI-3) protein ADI28460. The peptide was
CC identified in studies of RAI-3 phosphorylation in response to cigarette
CC smoke. Primers based on this sequence can be used in a PCR to identify
CC individuals at risk for developing chronic obstructive pulmonary disease
CC (COPD). RAI-3 is tyrosine phosphorylated and/or is associated/complexed
CC with tyrosine phosphorylated proteins only in cells that have been
CC exposed to cigarette smoke. Since RAI-3 is primarily expressed in lung
CC tissue, and since cigarette smoke is a major causative factor of COPD,
CC RAI-3 provides a novel cellular target for identifying modulators, e.g.
CC agonists or antagonists, useful for the treatment and/or prevention of
CC COPD and related disorders such as emphysema and chronic bronchitis. RAI-
CC 3 modulators, e.g. agonists and antagonists, especially antisense
CC compounds, can be used to treat COPD and other disorders an diseases
CC associated with regulation of NF-kB and/or its associated or interacting
CC signaling molecules. Single nucleotide polymorphisms (SNPs) detected in
CC the RAI-3 gene are useful for determining COPD association in
CC individuals. RAI-3 nucleic acid molecules and polypeptides are useful for
CC preventing, treating or ameliorating disorders related to aberrant GPCR
CC signaling or cell cycle regulation, pulmonary disorders, inflammatory
CC lung disorders, COPD, the underlying symptoms of COPD, COPD-related
CC disorders or conditions, autoimmune disorders, disorders related to
CC hyperimmune activity, inflammatory conditions, disorders related to
CC aberrant acute phase responses, hypercongenital conditions, birth
CC defects, necrotic lesions, wounds, organ transplant rejection, renal
CC diseases, ischaemia-reperfusion injury, heart disorders, disorders
CC related to aberrant signal transduction, proliferation disorders,
CC cancers, HIV infection, asthma, cystic fibrosis, pulmonary fibrosis,
CC ulcerative colitis, cerebral infarct, myocardial infarct, diabetic
CC neuropathy, allergic rhinitis, Crohn's disease, atherosclerosis,
CC rheumatoid arthritis, inflammatory/autoimmune disorders, glioblastoma,
CC pulmonary small cell undifferentiated carcinoma, carcinoma of the breast,
CC colon, lung, ovary, pancreas, prostate, non-Hodgkin's lymphoma, disorders
CC associated with aberrant cell adhesion, I-CAM function and/or regulation,
CC E-selectin function and/or regulation, or aberrant NF-kB function and/or
CC regulation (all claimed).
XX Sequence 40 BP; 11 A; 12 C; 10 G; 7 T; 0 U; 0 Other;
SQ Query Match 100.0%; Score 17; DB 12; Length 40;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCCCAGCTTGCCGAG 17
DB 1 GCCCAGCTTGCCGAG 17
RESULT 3
ADT96418
ID ADT96418 standard; cDNA; 497 BP.
XX
AC ADT96418;
```

```
XX 16-DEC-2004 (first entry)
DE Colon cancer associated human cDNA sequence #1925.
XX Colon cancer; T cell; tumour protein; C634S; C635S; C637S; C640S; C636S;
KW humoral immune response; cellular immune response; cytostatic;
KW immunostimulant; human; ss.
XX Homo sapiens.
OS US2003087818-A1.
XX 08-MAY-2003.
PD 01-FEB-2002; 2002US-00066543.
PF 02-FEB-2001; 2001US-0267400P.
XX 07-FEB-2001; 2001US-0267382P.
PR 11-MAY-2001; 2001US-0290322P.
PR 12-JUL-2001; 2001US-0305285P.
PR 16-AUG-2001; 2001US-0313077P.
XX (CORI-) CORIXA CORP.
PA Jiang Y, Chenault RA, Xu J, Indirias CY, Lodes MJ, Secrist H;
PI Carter D, Fanger GR, Smith CL, Durham M, Stolk JA;
XX WPI: 2003-040540/03.
DR New isolated nucleic acids and polypeptides capable of eliciting a
PT humoral and/or cellular immune response, useful for diagnosing,
PT preventing or treating cancer, particularly colon cancer.
XX Claim 1; SEQ ID NO 1937; 87pp; English.
PS The invention relates to polynucleotide and polypeptide sequences
XX associated with cancer, particularly colon cancer. Also disclosed are (i)
CC an expression vector comprising the polynucleotide, (ii) a host cell
CC transformed or transfected with the expression vector, (iii) an isolated
CC antibody, or its antigen-binding fragment, which specifically binds to
CC the polypeptide, (iv) a method of detecting or determining the presence
CC of cancer in a patient, (v) a fusion protein comprising at least one of
CC the polypeptides, (vi) an oligonucleotide that hybridises to the
CC polynucleotide sequence under highly stringent conditions, and (vii) a
CC method of stimulating and/or expanding T cells specific for a tumour
CC protein. The polypeptide specifically comprises the amino acid sequence
CC of C634S, C635S, C637S, C640S, C636S or one of the potential open reading
CC frames (ORFs) of C636S. These polypeptides are encoded by the
CC polynucleotide sequences, where both are capable of eliciting a humoral
CC and/or cellular immune response. The polynucleotides, polypeptides, and
CC antibodies are useful for diagnosing, preventing or treating cancer,
CC particularly colon cancer. The polynucleotide and polypeptide sequences
CC are also useful in DNA strand invasion, antisense inhibition, mutational
CC analysis, nucleic acid purification, isolation of transcriptionally
CC active genes, blocking or transcription factor binding, genome cleavage
CC or in situ hybridisation, and as enhancers of transcription or
CC biomarkers. This sequence represents a human colon cancer associated
CC cDNA. Note: The sequence data for this patent was obtained in electronic
XX format directly from the USPTO web site at seqdata.uspto.gov
SQ Sequence 497 BP; 126 A; 130 C; 129 G; 112 T; 0 U; 0 Other;
Query Match 100.0%; Score 17; DB 11; Length 497;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCCCAGCTTGCCGAG 17
DB 344 GCCCAGCTTGCCGAG 360
RESULT 4
```

```
ADX42900
ID ADX42900 standard; cDNA; 497 BP.
XX
AC ADX42900;
XX
DT 21-APR-2005 (first entry)
XX
DE Human cDNA encoding colon cancer protein SEQ ID NO 1937.
XX
XX Cytostatic; Immunostimulant; therapy; diagnosis; colon cancer; neoplasm;
XX ss; gene.
XX
OS Homo sapiens.
XX
FN WO200274156-A2.
XX
XX 26-SEP-2002.
XX
XX 01-FEB-2002; 2002WO-US002870.
XX
XX 02-FEB-2001; 2001US-0267400P.
XX
XX 07-FEB-2001; 2001US-0267382P.
XX
XX 11-MAY-2001; 2001US-0290322P.
XX
XX 12-JUL-2001; 2001US-0305265P.
XX
XX 16-AUG-2001; 2001US-0313077P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Jiang Y, Chenault RA, Xu J, Indirias CY, Lodes MJ, Secríst H;
XX Carter D, Fanger GR, Smith CL, Durham M, Stolk JA;
XX WPI; 2003-040540/03.
XX
XX New isolated nucleic acids and polypeptides capable of eliciting a
XX humoral and/or cellular immune response, useful for diagnosing,
XX preventing or treating cancer, particularly colon cancer.
XX
XX Claim 1; SEQ ID NO 1937; 244pp; English.
XX
XX The invention relates to a new isolated nucleic acid. The nucleic acids,
XX polypeptides, antibodies are useful for diagnosing, preventing or
XX treating cancer, particularly colon cancer. The nucleic acid and
XX polypeptides are also useful in DNA strand invasion, antisense
XX inhibition, mutational analysis, nucleic acid purification, isolation of
XX transcriptionally active genes, blocking or transcription factor binding,
XX genome cleavage or in situ hybridization, and as enhancers of
XX transcription or biomarkers. The kits are useful for detecting antibody
XX binding. The present sequence represents a human cDNA encoding a colon
XX cancer protein.
XX
XX Sequence 497 BP; 126 A; 130 C; 129 G; 112 T; 0 U; 0 Other;
XX
XX Query Match 100.0%; Score 17; DB 11; Length 497;
XX Best Local Similarity 100.0%; Pred. No. 63;
XX Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 GCCCAGCGTTGCCGAG 17
XX |||||
XX 344 GCCCAGCGTTGCCGAG 360
XX
XX RESULT 5
XX ADT96292/c
XX ID ADT96292 standard; cDNA; 552 BP.
XX
XX AC ADT96292;
XX
XX 16-DEC-2004 (first entry)
XX
XX Colon cancer associated human cDNA sequence #1799.
XX
XX Colon cancer; T cell; tumour protein; C634S; C637S; C640S; C636S;
XX humoral immune response; cellular immune response; cytostatic;
XX
```

```
immunostimulant; human; ss.
KW
XX Homo sapiens.
XX
XX US2003087818-A1.
XX
XX 08-MAY-2003.
XX
XX 01-FEB-2002; 2002US-00066543.
XX
XX 02-FEB-2001; 2001US-0267400P.
XX
XX 07-FEB-2001; 2001US-0267382P.
XX
XX 11-MAY-2001; 2001US-0290322P.
XX
XX 12-JUL-2001; 2001US-0305265P.
XX
XX 16-AUG-2001; 2001US-0313077P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Jiang Y, Chenault RA, Xu J, Indirias CY, Lodes MJ, Secríst H;
XX Carter D, Fanger GR, Smith CL, Durham M, Stolk JA;
XX WPI; 2003-040540/03.
XX
XX New isolated nucleic acids and polypeptides capable of eliciting a
XX humoral and/or cellular immune response, useful for diagnosing,
XX preventing or treating cancer, particularly colon cancer.
XX
XX Claim 1; SEQ ID NO 1811; 87pp; English.
XX
XX The invention relates to polynucleotide and polypeptide sequences
XX associated with cancer, particularly colon cancer. Also disclosed are (i)
XX an expression vector comprising the polynucleotide, (ii) a host cell
XX transformed or transfected with the expression vector, (iii) an isolated
XX antibody, or its antigen-binding fragment, which specifically binds to
XX the polypeptide, (iv) a method of detecting or determining the presence
XX of cancer in a patient, (v) a fusion protein comprising at least one of
XX the polypeptides, (vi) an oligonucleotide that hybridises to the
XX polynucleotide sequence under highly stringent conditions, and (vii) a
XX method of stimulating and/or expanding T cells specific for a tumour
XX protein. The polypeptide specifically comprises the amino acid sequence
XX of C634S, C635S, C637S, C640S, C636S or one of the potential open reading
XX frames (ORFs) of C636S. These polypeptides are encoded by the
XX polynucleotide sequences, where both are capable of eliciting a humoral
XX and/or cellular immune response. The polynucleotides, polypeptides, and
XX antibodies are useful for diagnosing, preventing or treating cancer,
XX particularly colon cancer. The polynucleotide and polypeptide sequences
XX are also useful in DNA strand invasion, antisense inhibition, mutational
XX analysis, nucleic acid purification, isolation of transcriptionally
XX active genes, blocking or transcription factor binding, genome cleavage
XX or in situ hybridisation, and as enhancers of transcription or
XX biomarkers. This sequence represents a human colon cancer associated
XX cDNA. Note: The sequence data for this patent was obtained in electronic
XX format directly from the USPTO web site at seqdata.uspto.gov
XX
XX Sequence 552 BP; 126 A; 141 C; 148 G; 137 T; 0 U; 0 Other;
XX
XX Query Match 100.0%; Score 17; DB 11; Length 552;
XX Best Local Similarity 100.0%; Pred. No. 64;
XX Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 GCCCAGCGTTGCCGAG 17
XX |||||
XX 154 GCCCAGCGTTGCCGAG 138
XX
XX RESULT 6
XX ADX42774/c
XX ID ADX42774 standard; cDNA; 552 BP.
XX
XX AC ADX42774;
XX
XX 21-APR-2005 (first entry)
XX
XX
```

```
DE Human cDNA encoding colon cancer protein SEQ ID NO 1811.
XX Cytostatic; Immunostimulant; therapy; diagnosis; colon cancer; neoplasm;
KW ss; gene.
OS
XX Homo sapiens.
XX
XX WO200274156-A2.
XX
XX PD 26-SEP-2002.
XX
XX PF 01-FEB-2002; 2002WO-US002870.
XX
XX PR 02-FEB-2001; 2001US-0267400P.
XX
XX PR 07-FEB-2001; 2001US-0267382P.
XX
XX PR 11-MAY-2001; 2001US-0290322P.
XX
XX PR 12-JUL-2001; 2001US-0305265P.
XX
XX PR 16-AUG-2001; 2001US-0313077P.
XX
XX (CORI-) CORIXA CORP.
XX
XX PI Jiang Y, Chenault RA, Xu J, Indrias CY, Lodes MJ, Secretist H;
XX Carter D, Fanger GR, Smith CL, Durham M, Stolk JA;
XX
XX WPI; 2003-040540/03.
XX
XX New isolated nucleic acids and polypeptides capable of eliciting a
XX humoral and/or cellular immune response, useful for diagnosing,
XX preventing or treating cancer, particularly colon cancer.
XX
XX Claim 1; SEQ ID NO 1811; 244pp; English.
XX
XX The invention relates to a new isolated nucleic acid. The nucleic acids,
XX polypeptides, antibodies are useful for diagnosing, preventing or
XX treating cancer, particularly colon cancer. The nucleic acid and
XX polypeptides are also useful in DNA strand invasion, antisense
XX inhibition, mutational analysis, nucleic acid purification, antisense
XX transcriptionally active genes, blocking or transcription factor binding,
XX genome cleavage or in situ hybridization, and as enhancers of
XX transcription or biomarkers. The kits are useful for detecting antibody
XX binding. The present sequence represents a human cDNA encoding a colon
XX cancer protein.
XX
XX SQ Sequence 552 BP; 126 A; 141 C; 148 G; 137 T; 0 U; 0 Other;
XX
XX Query Match 100.0%; Score 17; DB 11; Length 552;
XX Best Local Similarity 100.0%; Pred. No. 64;
XX Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 GCCACGCTTGCCGAG 17
XX |||||
XX DB 154 GCCACGCTTGCCGAG 138
XX
XX RESULT 7
XX AAZ90046
XX ID AAZ90046 standard; cDNA; 603 BP.
XX
XX AC AAZ90046;
XX
XX XX 09-MAY-2000 (first entry)
XX
XX DE Hydrophobic domain containing protein clone HP10549 coding sequence.
XX
XX KW Hydrophobic domain; clone HP10549; nutritional supplement; SCID; HIV;
XX cell proliferation; immune stimulant; immune deficiency; tumour; pain;
XX rheumatoid arthritis; insulin dependent diabetes mellitus; fertility;
XX myasthenia gravis; haematopoiesis regulator; tissue growth; depression;
XX anti-inflammatory; infection; bodily characteristic; ss.
XX
XX OS Homo sapiens.
XX
XX KW WO200000506-A2.
XX
XX PN
```

---

```
XX PD 06-JAN-2000.
XX
XX PF 18-JUN-1999; 99WO-JP003242.
XX
XX PR 26-JUN-1998; 98JP-00180008.
XX
XX PA (SAGA ) SAGAMI CHEM RES CENT.
XX (PROT-) PROTEGENE INC.
XX
XX PI Kato S, Kimura T;
XX
XX DR WPI; 2000-160665/14.
XX
XX DR P-PSDB; AAY78809.
XX
XX PT Novel human proteins having hydrophobic domains used for research and
XX diagnostic purposes.
XX
XX PS Claim 3; Page 88; 117pp; English.
XX
XX This sequence represents the hydrophobic domain containing protein, clone
XX HP10549 coding region. The sequence is isolated from a human stomach
XX cancer cell line. The invention relates to human proteins with
XX hydrophobic domains, the DNA and the cDNA encoding them. The
XX polynucleotides and proteins are predicted to have biological activities
XX which make them suitable for treating, preventing or ameliorating medical
XX conditions in humans and animals. Suggested activities include
XX nutritional activity (nutritional source or supplement); cytokine and
XX cell proliferation/differentiation activity; immune stimulating (e.g. as
XX vaccines) or suppressing activity (e.g. to treat various immune
XX deficiencies such as SCIDS or HIV, connective tissue disease, systemic
XX lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary
XX inflammation, Guillain-Barre syndrome, myasthenia gravis, graft-versus-host disease
XX and autoimmune inflammatory eye disease, as well as asthma, allergies and
XX organ transplantation); haematopoiesis regulating activity (e.g. in
XX treatment of myeloid or lymphoid cell deficiencies); tissue growth
XX activity (e.g. wound healing and tissue repair, ulcers, burns,
XX periodontal disease); activin/inhibin activity; chemotactic/chemokinetic
XX activity; haemostatic and thrombolytic activity (e.g. treating
XX haemophilias); receptor/ligand activity; anti-inflammatory activity; and
XX tumour inhibition activity. The polynucleotides are also stated to be
XX useful for gene therapy. Other activities include inhibiting infections
XX caused by bacteria, fungi, viruses and other parasites (e.g. Hepatitis,
XX malaria); effecting bodily characteristics such as, e.g. weight, colour,
XX skin, effecting biorhythms or cardiac cycles; enhancing fertility;
XX treatment of depression; treatment of pain; hormonal or endocrine
XX activity. The polynucleotides may also be used for recombinant expression
XX of the protein
XX
XX SQ Sequence 603 BP; 133 A; 179 C; 137 G; 154 T; 0 U; 0 Other;
XX
XX Query Match 100.0%; Score 17; DB 3; Length 603;
XX Best Local Similarity 100.0%; Pred. No. 64;
XX Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 GCCACGCTTGCCGAG 17
XX |||||
XX DB 550 GCCACGCTTGCCGAG 566
XX
XX RESULT 8
XX ABQ58527/c
XX ID ABQ58527 standard; cDNA; 620 BP.
XX
XX AC ABQ58527;
XX
XX DT 02-AUG-2002 (first entry)
XX
XX DE Human colon cancer related nucleotide sequence SEQ ID NO:2222.
XX
XX KW Human; colon cancer; cancer; tissue profiling; forensic; mapping;
XX genetic analysis; diagnostic; antisense therapy; gene; ss.
XX
XX PN
```





XX Novel isolated polypeptide encoded by breast cancer marker gene, useful  
PT for diagnosing, staging, monitoring, prognosing and treating diseases  
PT associated with breast cancer.

XX Disclosure; Page 187-188; 725pp; English.

XX The invention relates to an isolated polypeptide encoded by a breast  
CC cancer marker gene comprising any of 1417 21-805 nucleotide sequences,  
CC given in the specification. The methods of the invention are useful for  
CC diagnosing patients having an identified breast mass or symptoms  
CC associated with breast cancer, to diagnose breast cancer or its  
CC precursors, and for monitoring the efficacy of treatment of a breast  
CC cancer patient (e.g. efficacy of chemotherapy). The methods are also  
CC useful for evaluating a patient before, after or during therapy, to  
CC evaluate the reduction in a tumour burden. The breast cancer marker gene  
CC proteins are useful as immunogens for raising antibodies, by immunising a  
CC mammal with a breast cancer marker protein. The marker proteins are  
CC useful as bait proteins in a two-hybrid or three-hybrid assay, to  
CC identify other proteins which bind to or interact with the marker  
CC proteins. The breast cancer marker genes are useful as surrogate marker  
CC disease states, in particular, breast cancers. The breast cancer marker  
CC genes are useful as pharmacodynamic marker genes. An antibody which  
CC selectively binds to a protein of a breast cancer marker gene is useful  
CC for treating cancers, particularly breast cancers. The host cell of the  
CC invention is useful for producing non-human transgenic animals. This  
CC polynucleotide sequence represents one of the breast cancer marker genes  
CC of the invention

XX Sequence 642 BP; 148 A; 160 C; 180 G; 142 T; 0 U; 12 Other;  
SQ

Query Match 100.0%; Score 17; DB 10; Length 642;  
Best Local Similarity 100.0%; Pred. No. 64;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GCCCAGCGTTGGCCGAG 17  
|||  
Db 184 GCCCAGCGTTGGCCGAG 168

RESULT 11  
AD030035  
ID AD030035 standard; cDNA; 1074 BP.

XX AD030035;  
AC

XX 29-JUL-2004 (first entry)  
DT

XX Human GPCR RA13 polynucleotide, SEQ ID NO:1137.  
DE

XX G protein-coupled receptor; GPCR; drug screening; diagnosis;  
KW transgenic mouse; neurological disorder; adrenal gland disorder;  
KW colon disorder; intestinal disorder; cardiovascular disorder;  
KW muscular disorder; blood disorder; immune disorder; bone disorder;  
KW joint disorder; metabolic disorder; nutritive disorder; cancer;  
KW kidney disorder; liver disorder; lung disorder; breast disorder;  
KW ovary disorder; uterus disorder; prostate disorder; testis disorder;  
KW skin disorder; stomach disorder; pancreas disorder; spleen disorder;  
KW thymus disorder; thyroid disorder; antiparkinsonian; antimanic;  
KW cytostatic; antiinflammatory; vasotropic; antiangiogenic; antidiabetic;  
KW CNS; central nervous system; respiratory; antidiarrhoeic; antidiabetic;  
KW virucide; hepatotropic; antibacterial; antianemic; antiseborrheic;  
KW dermatological; antiulcer; antihypertensive; antiallergic; anorectic;  
KW immunosuppressive; nephrotropic; gene therapy; GPCR modulator; human;  
KW gene; ss.

XX Homo sapiens.  
OS

XX W02004040000-A2.  
PN

XX 13-MAY-2004.  
PD

XX

PF 09-SEP-2003; 2003WO-US028226.  
XX  
PR 09-SEP-2002; 2002US-0409303p.  
PR 09-APR-2003; 2003US-0461329p.  
XX  
XX (PRIM-) PRIMAL INC.  
XX  
XX Gaitanaris GA, Bergmann JE, Gragerov A, Hohmann J, Li F;  
PI Madisen L, McIlwain KL, Pavlova MN, Vassilatis D, Zeng H;  
XX  
XX WPI; 2004-390329/36.  
DR P-PSDB; ADO29632.  
XX  
XX Novel mammalian G protein coupled receptors, useful for identifying  
PT compounds that modulates diagnosing and treating disease condition  
PT associated with GPCR dysfunction e.g. autoimmune diseases, angina  
PT pectoris, Parkinson's disease.  
XX  
XX Claim 151; SEQ ID NO 1137; 542pp; English.  
PS  
CC The invention relates to human and mouse G protein-coupled receptors  
CC (GPCRs) and nucleic acids encoding them. The invention also relates to  
CC sequences at least 90% identical to the GPCR proteins and nucleic acids  
CC of the invention; methods of treating, preventing or diagnosing diseases  
CC associated with GPCRs of the invention; methods of screening for  
CC compounds useful in the treatment of GPCR-related diseases; a transgenic  
CC mouse comprising a GPCR gene of the invention; a mouse comprising a  
CC mutation in a GPCR transgene or in an endogenous GPCR gene; cells derived  
CC from the transgenic mice; kits comprising several mice, each of which has  
CC a mutation in a different GPCR gene of the invention; and kits comprising  
CC probes which hybridise to GPCR polynucleotides of the invention. The  
CC invention further discloses variants of the GPCR polypeptides and vectors  
CC comprising a GPCR nucleic acid. The GPCR nucleic acids and proteins may  
CC be used in the diagnosis, treatment or prevention of a wide variety of  
CC diseases including neurological disorders (e.g., Alzheimer's disease,  
CC depression, diabetic neuropathy, Parkinson's disease or schizophrenia);  
CC disorders of the adrenal gland; disorders of the colon or intestine  
CC (e.g., Crohn's disease, diarrhoea, food poisoning or irritable bowel  
CC syndrome); cardiovascular disorders (e.g., angina, cardiac arrhythmia or  
CC myocardial infarction); muscular disorders; blood disorders (e.g.,  
CC anaemia or leukaemia); immune disorders (e.g., autoimmune disorders or  
CC AIDS); bone and joint disorders (e.g., osteoarthritis, rheumatoid  
CC arthritis, gout or osteoporosis); metabolic or nutritive disorders (e.g.,  
CC obesity, enzyme deficiency-related diseases or vitamin deficiency-related  
CC diseases); and disorders of the kidney, liver, lung, breast, ovary,  
CC uterus, prostate, testis, skin, stomach, pancreas, spleen, thymus and  
CC thyroid (e.g., cancers). The present sequence represents a GPCR-encoding  
CC nucleic acid of the invention. Note: The full sequence data for this  
CC patent did not form part of the printed specification; those sequences  
CC not shown were obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 1074 BP; 215 A; 316 C; 259 G; 284 T; 0 U; 0 Other;  
SQ

Query Match 100.0%; Score 17; DB 12; Length 1074;  
Best Local Similarity 100.0%; Pred. No. 64;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GCCCAGCGTTGGCCGAG 17  
|||  
Db 1018 GCCCAGCGTTGGCCGAG 1034

RESULT 12  
AAZ61776  
ID AAZ61776 standard; cDNA; 1212 BP.  
XX  
XX AAZ61776;  
AC

XX 27-MAR-2000 (first entry)  
DT

XX cDNA encoding human skin cell transmembrane protein, SEQ ID NO:249.  
DE

XX

KW Skin, dermal papilla; keratinocyte; neonatal foreskin fibroblast;  
KW embryonic skin cell; keratinocyte stem cell; transit amplifying cell;  
KW secreted; transmembrane; inflammation; cancer; neurological disease;  
KW angiogenesis; tumour vascularisation; growth disorder;  
KW developmental disorder; skin wound; hair follicle disorder;  
KW anti-inflammatory; cytostatic; neuroprotective; vulnery; ss.  
XX  
OS Homo sapiens.  
XX  
XX WO9955865-A1.  
XX  
XX 04-NOV-1999.  
XX  
XX 29-APR-1999; 99WO-NZ000051.  
PF  
XX 29-APR-1998; 98US-00069726.  
PR  
XX 09-NOV-1998; 98US-00188930.  
PR  
XX (GENE-) GENESIS RES & DEV CORP LTD.  
PA  
XX Strachan L, Sleeman M, Watson JD, Onrust R, Kumble A, Murison JG;  
PI  
XX WPI; 2000-072177/06.  
XX  
XX Novel polynucleotides useful for the treatment of various conditions  
PT including wounds and cancer.  
PT  
XX Claim 1; Page 155; 235pp; English.  
XX  
XX The invention relates to novel nucleic acid sequences derived from rat  
CC dermal papilla, human keratinocytes and neonatal foreskin fibroblasts,  
CC and mouse embryonic skin, keratinocyte stem cells and transit amplifying  
CC cells. Polypeptides of the invention may be used to treat inflammation,  
CC cancer and neurological diseases. The proteins may be used to stimulate  
CC the growth and motility of keratinocytes, to inhibit the growth of cancer  
CC cells, to modulate angiogenesis and tumour vascularisation, to modulate  
CC skin inflammation, to modulate epithelial cell growth and to inhibit  
CC binding of HIV-1 to leukocytes. The invention may also be used to treat  
CC growth and developmental defects, skin wounds and hair follicle  
CC disorders. Sequences AAZ61606-261832 represent cDNA sequences derived  
CC from several mouse, rat or human skin cell types. Sequences AAZ61606-  
CC Z61649, AAZ61725-261765, AAZ61802-261811 and AAZ61826 encode proteins  
CC with an N-terminal signal sequence, indicating that the proteins are  
CC secreted. Sequences AAZ61650-261668, AAZ61766-261780, AAZ61812-261817 and  
CC AAZ61827-261829 encode proteins with one or more putative transmembrane  
CC domains  
XX  
SQ Sequence 1212 BP; 242 A; 350 C; 312 G; 308 T; 0 U; 0 Other;  
Query Match 100.0%; Score 17; DB 3; Length 1212;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GCCCAGCGTTGGCCGAG 17  
Db 990 GCCCAGCGTTGGCCGAG 1006  
RESULT 13  
AAC99709  
ID AAC99709 standard; cDNA; 1212 BP.  
XX  
XX AAC99709;  
AC  
XX 08-MAR-2001 (first entry)  
DT  
XX Skin cell cDNA, SEQ ID NO: 249.  
DE  
XX Human; skin cell; cytostatic; antiinflammatory; anti-HIV; nootropic;  
XX neuroprotective; vulnery; immunomodulatory; vaccine;  
KW keratinocyte growth stimulation; cancer; angiogenesis inhibition;  
KW inflammation; neurological disease; ss.  
XX

OS Homo sapiens.  
XX  
XX WO200069884-A2.  
PN  
XX 23-NOV-2000.  
PD  
XX  
XX 15-MAY-2000; 2000WO-NZ000075.  
PF  
XX  
XX 14-MAY-1999; 99US-00312283.  
PR  
XX (GENE-) GENESIS RES & DEV CORP LTD.  
PA  
XX Watson JD, Strachan L, Onrust R, Sleeman M, Kumble KD;  
PI Murison JG;  
PI  
XX WPI; 2001-007495/01.  
DR  
XX P-PSDB; AAB55958.  
DR  
XX New isolated polynucleotide used in the identification of genetic  
PT disorders and encoding polypeptides used for treating inflammatory  
PT disease, cancer and neurological diseases.  
PT  
XX Claim 1; Page 210-211; 352pp; English.  
XX  
XX The present polynucleotide encodes a polypeptide which is expressed in  
CC mammalian skin cells. The polypeptide is useful for stimulating  
CC keratinocyte growth and motility, inhibiting the growth of cancer cells,  
CC modulating angiogenesis, inhibiting angiogenesis and vascularisation of  
CC tumours, modulating skin inflammation, stimulating the growth of  
CC epithelial cells, inhibiting the binding of human immunodeficiency virus  
CC (HIV)-1 to leukocytes, and treating inflammatory disease, cancer and  
CC neurological diseases. The polynucleotide can be used as a marker, in the  
CC identification of genetic disorders, and for the design of  
CC oligonucleotides for examining expression patterns  
XX  
SQ Sequence 1212 BP; 242 A; 350 C; 312 G; 308 T; 0 U; 0 Other;  
Query Match 100.0%; Score 17; DB 4; Length 1212;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GCCCAGCGTTGGCCGAG 17  
Db 990 GCCCAGCGTTGGCCGAG 1006  
RESULT 14  
ABL34861  
ID ABL34861 standard; cDNA; 1212 BP.  
XX  
XX ABL34861;  
AC  
XX 04-APR-2002 (first entry)  
DT  
XX Human cDNA isolated from skin cells SEQ ID NO: 249.  
DE  
XX Human; rat; mouse; skin cell; skin wound; cancer; growth defect;  
XX developmental defect; inflammatory disease; dermatological; vulnery;  
KW immunomodulator; anti-inflammatory; cytostatic; neuroprotective; gene;  
KW ss.  
XX  
XX Homo sapiens.  
OS  
XX WO200190357-A1.  
PN  
XX 29-NOV-2001.  
PD  
XX  
XX 24-MAY-2001; 2001WO-NZ000099.  
PF  
XX 24-MAY-2000; 2000US-0206650P.  
PR  
XX 25-JUL-2000; 2000US-0221232P.  
PR  
XX (GENE-) GENESIS RES & DEV CORP LTD.  
PA

XX Watson JD, Strachan L, Sleeman M, Onrust R, Murison JG;  
PI Kumble KD;  
XX WPI; 2002-122020/16.  
XX New polynucleotides and polypeptides encoded by the polynucleotides  
PT isolated from skin cells, useful for treating skin wounds, cancers,  
PT growth and developmental defects, inflammatory diseases, or for  
PT modulating immune responses.  
XX Claim 1; Page 175; 466pp; English.  
XX The present invention provides the protein and coding sequences of cDNAs  
CC isolated from human, murine and rat skin cell libraries. The sequences  
CC can be used in the development of therapeutic agents useful in the  
CC treatment of skin diseases, including skin wounds, cancer, growth  
CC defects, developmental defects and inflammatory diseases. The proteins  
CC have important roles in the induction of hair growth, cell proliferation  
CC and cell-cell interaction, in maintaining tissue integrity, in wound  
CC healing and in modulating immune responses. The present sequence is a  
CC cDNA of the invention  
XX  
SQ Sequence 1212 BP; 242 A; 350 C; 312 G; 308 T; 0 U; 0 Other;  
Query Match 100.0%; Score 17; DB 6; Length 1212;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GCCCAGCTTGCCGAG 17  
Db 990 GCCCAGCTTGCCGAG 1006  
RESULT 15  
ACA04775  
ID ACA04775 standard; cDNA; 1228 BP.  
AC ACA04775;  
XX  
XX 28-MAY-2003 (first entry)  
XX cDNA encoding human membrane associated protein fragment #223.  
XX Human; ss; gene; microarray; membrane-associated protein; neuropathology;  
KW immunopathology; pancreatic disease; cancer; diabetes; hyperlipidaemia;  
KW pancreatic cholera; Alzheimer's disease; Huntington's disease; sarcoma;  
KW fibrocystic disease; leukaemia; adenocarcinoma; AIDS; allergy; anaemia;  
KW asthma; gout; dementia.  
XX Homo sapiens.  
XX US6492505-B1.  
XX 10-DEC-2002.  
XX 31-JAN-2000; 2000US-00495050.  
XX 01-FEB-1999; 99US-0118318P.  
XX (INCYTE) INCYTE GENOMICS INC.  
XX Reddy R, Guegler KJ, Au-Young J;  
XX WPI; 2003-327324/31.  
XX Combination for research/diagnostic applications and for monitoring  
PT treatment of e.g., cancer, comprises polynucleotides comprising a  
PT fragment of gene encoding membrane-associated proteins, receptors or ion  
PT channels.  
XX Claim 1; Col 215-216; 147pp; English.

CC The invention relates to a combination comprising several polynucleotide  
CC sequences comprising a fragment of gene encoding membrane-associated  
CC proteins, receptors or ion channels. The combination is useful as a  
CC probe, for research and diagnostic applications, for monitoring the  
CC expression of several expressed polynucleotides, in the diagnosis and  
CC monitoring of treatment of pancreatic disease, cancer, immunopathology or  
CC neuropathology, for investigating an individual's predisposition to the  
CC above disease, in genetic or gene expression analysis of polynucleotide  
CC sequences, to investigate cellular responses to infection or drug  
CC treatment, as hybridisable array elements in a microarray, to purify a  
CC subpopulation of mRNAs, cDNAs or genomic fragments in a sample, in  
CC diagnostics, prognostics and treatment regimens, in drug discovery and  
CC development, in toxicological and carcinogenicity studies, and in  
CC forensics or pharmacogenomics, to monitor the progression of disease, to  
CC monitor the efficacy of treatment, to diagnose the conditions of the  
CC pancreas e.g. diabetes, pancreatic cholera, hyperlipidaemia or  
CC fibrocystic disease, to diagnose a cancer e.g. leukaemia, adenocarcinoma  
CC or sarcoma, to diagnose immunopathologies e.g. AIDS, allergies, anaemia,  
CC asthma or gout, to diagnose neuropathologies e.g. Alzheimer's disease,  
CC dementia or Huntington's disease, to rapidly screen large numbers of  
CC candidate drug molecules and as query sequences against GenBank.  
CC SwissProt, BLOCKS and PRINTS databases. The combination is employed to  
CC fine tune the treatment regimen and thus the expression patterns  
CC associated with undesirable side effects are avoided. The present  
CC sequence represents a cDNA encoding a fragment of gene encoding human  
CC membrane-associated proteins, receptors or ion channels  
XX  
SQ Sequence 1228 BP; 272 A; 341 C; 283 G; 332 T; 0 U; 0 Other;  
Query Match 100.0%; Score 17; DB 8; Length 1228;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GCCCAGCTTGCCGAG 17  
Db 541 GCCCAGCTTGCCGAG 557  
RESULT 16  
AEA00112  
ID AEA00112 standard; cDNA; 1460 BP.  
XX  
XX AEA00112;  
XX  
XX 28-JUL-2005 (first entry)  
XX Human TAT143 cDNA SEQ ID NO:64.  
XX ss; gene; tumor-associated antigen; cytostatic; breast tumor;  
KW endocrine disease; gynecology and obstetrics; neoplasm; colon tumor;  
KW gastrointestinal disease; rectal tumor; endometroid carcinoma;  
KW genitourinary disease; renal tumor; lung tumor; respiratory disease;  
KW ovary tumor; skin tumor; liver tumor.  
XX Homo sapiens.  
XX  
XX Key Location/Qualifiers  
FT CDS 128..1201  
FT /\*tag= a  
FT /product= "TAT143"  
XX  
XX US2005106644-A1.  
XX 19-MAY-2005.  
XX  
XX 08-SEP-2004; 2004US-00936626.  
XX 20-JUN-2001; 2001US-0299500P.  
XX 29-JUN-2001; 2001US-0301880P.  
XX 18-SEP-2001; 2001US-0323288P.  
XX 19-JUN-2002; 2002US-0017488.  
XX 26-MAR-2004; 2004US-0557116P.  
XX 04-AUG-2004; 2004US-0598899P.

```
XX PA (GETH ) GENENTECH INC.
XX PI Cairns B, Chen R, Frantz G, Hillan KJ, Koeppen H, Phillips HS;
XX PI Polakis P, Spencer SD, Smith V, Williams PM, Wu TD, Zhang Z;
XX PI Sliwkowski M;
XX XX
DR WPI; 2005-384304/39.
DR P-PSDB; AEA00190.
XX XX
XX PT Novel isolated antibody capable of binding to tumor-associated antigenic
XX PT target polypeptide, useful for treating cell proliferative disorder e.g.
XX PT cancer.
XX XX
PS Claim 1; SEQ ID NO 64; 337pp; English.
XX XX
XX CC The invention relates to a novel isolated antibody binding to a
XX CC polypeptide having at least 80% sequence identity to a polypeptide having
XX CC any one of 76 fully defined 182-910 amino acid tumor-associated antigenic
XX CC target polypeptide (TAT) sequences (AEA00127-AEA00202) given in the
XX CC specification, a polypeptide having any one of (AEA00127-AEA00202),
XX CC lacking its associated signal peptide, or an extracellular domain of a
XX CC polypeptide having any one of (AEA00127-AEA00202). The polypeptide is
XX CC encoded by the nucleotide sequence having any one of (AEA00049-AEA00126).
XX CC An antibody of the invention has cytostatic activity. The antibody is
XX CC useful for inhibiting growth of a cell expressing TAT188, which involves
XX CC contacting the cell with the antibody. The cell is a cancer cell chosen
XX CC from breast, colon, rectum, endometrium, kidney, lung, ovary, skin and
XX CC liver cell. The cancer cell is a mammalian cell, preferably a human cell.
XX CC The antibody is also useful for detecting the level of TAT188
XX CC polypeptide expressed in a test cell relative to a control cell, and for
XX CC detecting the level of TAT188 polypeptide or a polypeptide having at
XX CC least 80 % sequence identity to the TAT188 polypeptide sequence in a test
XX CC cell relative to a control cell. The antibody is useful for treating a
XX CC preventing a cell proliferative disorder associated with increased
XX CC expression or activity of a polypeptide having at least 80 % identity to
XX CC a TAT188 polypeptide sequence. The cell proliferative disorder is cancer.
XX CC The method of the invention is useful for inhibiting the growth of a
XX CC cancer cell. The present sequence encodes a polypeptide of the invention.
XX XX
SQ Sequence 1460 BP; 302 A; 418 C; 362 G; 378 T; 0 U; 0 Other;
Query Match 100.0%; Score 17; DB 14; Length 1460;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCCCAGCGTTGGCGGAG 17
Db 1145 GCCCAGCGTTGGCGGAG 1161
RESULT 17
AEA00632
ID AEA00632 standard; cDNA; 1460 BP.
XX AC AEA00632;
XX XX
XX DT 28-JUL-2005 (first entry)
XX XX
XX DE Human TAT143 cDNA sequence SeqID64.
XX XX
XX KW antibody identification; tumor-associated antigen; cytostatic;
XX KW RNA interference; gene therapy; cell death; cancer; breast tumor;
XX KW colon tumor; rectal tumor; renal tumor; lung tumor; ovary tumor;
XX KW skin tumor; liver tumor; gene; ss; TAT143.
XX OS Homo sapiens.
XX XX
XX PN US2005107595-A1.
XX XX
XX PD 19-MAY-2005.
XX XX
XX PF 10-SEP-2004; 2004US-00938061.
```

```
XX PR 20-JUN-2001; 2001US-0299500P.
XX PR 29-JUN-2001; 2001US-0301880P.
XX PR 18-SEP-2001; 2001US-0323268P.
XX PR 19-JUN-2002; 2002US-00177488.
XX PR 26-MAR-2004; 2004US-0557116P.
XX PR 04-AUG-2004; 2004US-0598899P.
XX PA (GETH ) GENENTECH INC.
XX XX
XX PI Cairns B, Chen R, Frantz G, Hillan KJ, Koeppen H, Phillips HS;
XX PI Polakis P, Spencer SD, Smith V, Williams PM, Wu TD, Zhang Z;
XX PI Sakanaka C, Chuntharapai A, Reed CJ;
XX XX
XX DR WPI; 2005-371577/38.
XX DR P-PSDB; AEA00710.
XX XX
XX PT Novel isolated antibody e.g. anti-E16 or anti-TAT112 antibody that binds
XX PT to tumor-associated antigenic target polypeptide, useful for diagnosing
XX PT or treating cancer.
XX XX
XX PS Claim 1; SEQ ID NO 64; 96pp; English.
XX XX
XX CC This invention relates to a novel isolated antibody, for example anti-E16
XX CC or anti-TAT112 antibody, that binds to a tumor-associated antigenic
XX CC target polypeptide (TAT) and that lacks an associated signal peptide
XX CC sequence. The invention may be useful for the development of compounds
XX CC with a cytostatic activity acting as antagonists of the TAT188
XX CC polypeptide or RNA interference whilst the disclosed sequences may be
XX CC useful for gene therapy. The invention is useful for inducing the death
XX CC of a cell (such as a cancer cell chosen from breast, colon, rectum,
XX CC endometrium, kidney, lung, ovary, skin and liver) to which it binds,
XX CC inhibiting proliferation or promoting cell death of a cell expressing
XX CC TAT188. In addition, the invention may be useful for detecting the level
XX CC of TAT188 polypeptide in a test cell relative to a control cell, or
XX CC treating or preventing a cell proliferative disorder associated with
XX CC increased expression of TAT188. The novel antibody of the invention is
XX CC useful for inhibiting the growth of a cancer cell and may be useful for
XX CC diagnosing or treating cancer. The present sequence is that of the human
XX CC TAT143 cDNA which encodes a protein against which an antibody of the
XX CC invention may be targeted.
XX XX
SQ Sequence 1460 BP; 302 A; 418 C; 362 G; 378 T; 0 U; 0 Other;
Query Match 100.0%; Score 17; DB 14; Length 1460;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCCCAGCGTTGGCGGAG 17
Db 1145 GCCCAGCGTTGGCGGAG 1161
RESULT 18
AAF58615
ID AAF58615 standard; cDNA; 1619 BP.
XX AC AAF58615;
XX XX
XX DT 24-APR-2001 (first entry)
XX XX
XX DE Human RECAP polynucleotide, SEQ ID NO: 43.
XX KW Human; RECAP; receptors and associated proteins; cerebroprotective;
XX KW neurotropic; neuroprotective; anticonvulsant; antiparkinsonian; anti-HIV;
XX KW antidiabetic; immunostimulant; immunomodulator; antiinflammatory;
XX KW antithyroid; immunosuppressive; nephrotropic; antigen; thyromimetic;
XX KW cytostatic; antibacterial; virucide; fungicide; protozoacide; cancer; ss.
XX KW antiarteriosclerotic; hepatotropic; gene therapy; infection; cancer; ss.
XX OS Homo sapiens.
XX XX
XX PN WO200107612-A2.
```

```

XX PD 01-FEB-2001.
XX PF
XX PR 21-JUL-2000; 2000WO-US020035.
XX XX
XX PR 21-JUL-1999; 99US-0145232P.
XX PR 07-OCT-1999; 99US-0158578P.
XX PR 12-NOV-1999; 99US-0165192P.
XX XX
XX PA (INCY-) INCYTE GENOMICS INC.
XX XX
XX PI Au-Young J, Bandman O, Tang YT, Yue H, Azimzai Y, Burford N;
XX PI Baughn MR, Lu DAM, Hillman JL, Patterson C, Lal P;
XX DR WPI; 2001-168554/17.
XX DR P-PSDB; AAB68891.
XX XX
XX PT Novel receptors and associated proteins for diagnosis and treatment of
XX PT neurological disorders, immunological disorders including autoimmune/
XX PT inflammatory disorders and cell proliferative disorders such as cancer.
XX PS Example 5; Page 126-127; 128pp; English.
XX CC The present sequence encodes a human RECAP (receptors and associated
XX CC proteins) polypeptide. RECAP polynucleotides and polypeptides are useful
XX CC in the diagnosis, treatment and prevention of neurological disorders such
XX CC as stroke, Alzheimer's disease, Pick's disease, Huntington's disease,
XX CC dementia, Parkinson's disease, Down's syndrome, anyotrophic lateral
XX CC sclerosis, multiple sclerosis, bacterial and viral meningitis, CJD
XX CC (Creutzfeldt-Jakob disease), GSS (Gerstmann -Straussler-Scheinker
XX CC syndrome); immunological disorders, including autoimmune/inflammatory
XX CC disorders such as AIDS, DiGeorge's syndrome, severe combined
XX CC immunodeficiency disease (SCID), Chediak-Higashi syndrome, Crohn's disease,
XX CC disease, Addison's disease, autoimmune thyroiditis, Crohn's disease,
XX CC diabetes mellitus, Good pasture's syndrome, gout, Grave's disease,
XX CC Hashimoto's thyroiditis, Sjogren's syndrome, Werner's syndrome, viral,
XX CC bacterial, fungal, parasitic, protozoal, and helminthic infections; and
XX CC cell proliferation disorders such as arteriosclerosis, atherosclerosis,
XX CC cirrhosis, hepatitis and cancer
XX XX
XX SQ Sequence 1619 BP; 331 A; 463 C; 400 G; 425 T; 0 U; 0 Other;
Query Match 100.0%; Score 17; DB 4; Length 1619;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACGCTTGCCGAG 17
DB 1137 GCCACGCTTGCCGAG 1153

RESULT 19
AAZ90056
ID AAZ90056 standard; cDNA; 1718 BP.
XX AC
XX AC AAZ90056;
XX XX
XX DT 09-MAY-2000 (first entry)
XX XX
XX DE Hydrophobic domain containing protein clone HP10549 nucleotide sequence.
XX KW Hydrophobic domain; clone HP10549; nutritional supplement; SCID; HIV;
XX KW cell proliferation; immune stimulant; immune deficiency; tumour; pain;
XX KW rheumatoid arthritis; insulin dependent diabetes mellitus; fertility;
XX KW myasthenia gravis; haematopoiesis regulator; tissue growth; depression;
XX KW anti-inflammatory; infection; bodily characteristic; ss.
XX OS Homo sapiens.
XX XX
XX PN W0200000506-A2.
XX XX
XX PD 06-JAN-2000.
XX XX

```

```

PF 18-JUN-1999; 99WO-JP003242.
XX XX
XX PR 26-JUN-1998; 98JP-00180008.
XX XX
XX PA (SAGA ) SAGAMI CHEM RES CENT.
XX PA (PROT-) PROTEGENE INC.
XX PI Kato S, Kimura T;
XX XX
XX DR WPI; 2000-160665/14.
XX DR P-PSDB; AAY78809.
XX XX
XX PT Novel human proteins having hydrophobic domains used for research and
XX PT diagnostic purposes.
XX PS Claim 4; Page 111-113; 117pp; English.
XX CC This sequence represents the hydrophobic domain containing protein, clone
XX CC HP10549 nucleotide sequence. The sequence is isolated from a human
XX CC stomach cancer cell line. The invention relates to human proteins with
XX CC hydrophobic domains, the DNA and the cDNA encoding them. The
XX CC polynucleotides and proteins are predicted to have biological activities
XX CC which make them suitable for treating, preventing or ameliorating medical
XX CC conditions in humans and animals. Suggested activities include
XX CC nutritional activity (nutritional source or supplement); cytokine and
XX CC cell proliferation/differentiation activity; immune stimulating (e.g. as
XX CC vaccines) or suppressing activity (e.g. to treat various immune
XX CC deficiencies such as SCIDS or HIV, connective tissue disease, systemic
XX CC lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary
XX CC inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin
XX CC dependent diabetes mellitus, myasthenia gravis, graft-versus-host disease
XX CC and autoimmune inflammatory eye disease, as well as asthma, allergies and
XX CC organ transplantation); haematopoiesis regulating activity (e.g. in
XX CC treatment of myeloid or lymphoid cell deficiencies); tissue growth
XX CC activity (e.g. wound healing and tissue repair, ulcers, burns,
XX CC periodontal disease); activin/inhibin activity; chemotactic/chemokinetic
XX CC activity; haemostatic and thrombolytic activity (e.g. treating
XX CC haemophilias); receptor/ligand activity; anti-inflammatory activity; and
XX CC tumour inhibition activity. The polynucleotides are also stated to be
XX CC useful for gene therapy. Other activities include inhibiting infections
XX CC caused by bacteria, fungi, viruses and other parasites (e.g. Hepatitis,
XX CC malaria); effecting bodily characteristics such as, e.g. weight, colour,
XX CC skin; effecting biorhythms or cardiac cycles; enhancing fertility;
XX CC treatment of depression; treatment of pain; hormonal or endocrine
XX CC activity. The polynucleotides may also be used for recombinant expression
XX CC of the protein
XX XX
XX SQ Sequence 1718 BP; 381 A; 495 C; 377 G; 465 T; 0 U; 0 Other;
Query Match 100.0%; Score 17; DB 3; Length 1718;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACGCTTGCCGAG 17
DB 561 GCCACGCTTGCCGAG 577

RESULT 20
ADF70573
ID ADF70573 standard; DNA; 1788 BP.
XX AC
XX AC ADF70573;
XX XX
XX DT 12-FEB-2004 (first entry)
XX XX
XX DE Orphan receptor ligand-related human protein gene SeqID196.
XX KW ligand; orphan receptor protein; fusion protein; fluorescent protein;
XX KW cell expression; green fluorescent protein; GFP; GFP-1; wild-type GFP;
XX KW GFPuv; Enhanced GFP; EGFP; human; gene; ds.
XX OS Homo sapiens.

```

```
XX PN WO2003071272-A1.
XX PD 28-AUG-2003.
XX PX 21-FEB-2003; 2003WO-JP001901.
XX PR 22-FEB-2002; 2002JP-00045728.
XX PR 23-JUL-2002; 2002JP-00213949.
XX PR 11-OCT-2002; 2002JP-00298237.
XX PA (TAKE ) TAKEDA CHEM IND LTD.
XX PI Hinuma S, Fujii R, Ogi K, Komatsu H, Kawamata Y, Hosoya M;
XX PX WPI; 2003-697654/66.
XX DR P-PSDB; ADF70471.
XX DR
XX PX Transformation of cells with a fusion protein of an orphan receptor
XX PT protein with a fluorescent protein useful for identification of ligands
XX PT to the orphan receptor.
XX PX Example 4; SEQ ID NO 196; 594pp; Japanese.
XX CC This invention relates to a novel method of identifying ligands to an
XX CC orphan receptor protein which comprises transforming cells with DNA
XX CC encoding a fusion protein of the orphan receptor with a fluorescent
XX CC protein, so that the fusion protein is expressed in the cells (or cell
XX CC membranes isolated from them) and contacting the cells with the potential
XX CC ligand to be tested. A suitable fluorescent protein for incorporation in
XX CC the fusion protein is green fluorescent protein (GFP), for example GFP-1,
XX CC wild-type GFP, GFPuv or Enhanced GFP (EGFP). The method is useful for the
XX CC identification of ligands binding to an orphan receptor protein.
XX PX Sequence 1788 BP; 444 A; 473 C; 412 G; 459 T; 0 U; 0 Other;
XX
XX Query Match 100.0%; Score 17; DB 10; Length 1788;
XX Best Local Similarity 100.0%; Pred. No. 65;
XX Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
Oy 1 GCCCAGCGCTTGCCGAG 17
Db 1018 GCCCAGCGCTTGCCGAG 1034
| | | | | | | | | | | | | | | | | |
RESULT 21
ABT10173
ID ABT10173 standard; cDNA; 2302 BP.
XX AC ABT10173;
XX DT 04-DEC-2002 (first entry)
XX DE Human breast cancer associated coding sequence SEQ ID NO: 307.
XX PX Human; breast specific gene; breast cancer; differential expression;
XX KW cytostatic; gene therapy; gene; ss.
XX OS Homo sapiens.
XX PN WO200259271-A2.
XX PD 01-AUG-2002.
XX PX 25-JAN-2002; 2002WO-US002176.
XX PR 25-JAN-2001; 2001US-0263757P.
XX PR 25-APR-2001; 2001US-0286090P.
XX PR 23-MAY-2001; 2001US-0292517P.
XX PX (GENE-) GENE LOGIC INC.
XX PI Orr MS, Nation M, Diggins JC, Zeng W;
```

```
XX WI; 2002-674803/72.
XX DR
XX PX Diagnosing breast cancer in a patient comprises detecting the level of
XX PT gene expression in cell or tissue samples, where a differential gene
XX PT expression is indicative of breast cancer.
XX PS Claim 1; SEQ ID NO 307; 260pp + Sequence Listing; English.
XX PX The present invention relates to methods of diagnosing breast cancer in a
XX CC patient, which comprise detecting the level of expression in a tissue
XX CC sample of two or more genes selected from those shown in ABT09867-
XX CC AB11112, where a differential expression of the genes indicates breast
XX CC cancer. The methods are useful in diagnosing, treating, detecting the
XX CC progression, and in monitoring treatment of breast cancer in patients.
XX CC The methods are also useful as a screening tool for agents that modulate
XX CC the onset or progression of breast cancer. The breast cancer genes may be
XX CC used as diagnostic markers for the prediction or identification of the
XX CC malignant state of breast tissue, for confirming the type and progression
XX CC of cancer, and for drug screening and assays. The present sequence is a
XX CC coding sequence of the invention. Note: The sequence data for this patent
XX CC did not form part of the printed specification, but was obtained in
XX CC electronic format directly from WIPO at
XX CC ftp.wipo.int/pub.published_pct_sequences
XX PX Sequence 2302 BP; 494 A; 666 C; 533 G; 609 T; 0 U; 0 Other;
XX
XX Query Match 100.0%; Score 17; DB 6; Length 2302;
XX Best Local Similarity 100.0%; Pred. No. 66;
XX Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
Oy 1 GCCCAGCGCTTGCCGAG 17
Db 1117 GCCCAGCGCTTGCCGAG 1133
| | | | | | | | | | | | | | | | | |
RESULT 22
ACC58386
ID ACC58386 standard; cDNA; 2302 BP.
XX AC ACC58386;
XX DT 26-AUG-2003 (first entry)
XX DE Human GPCR-like retinoic acid-induced gene 1.
XX PX Human; retinoic acid-induced gene 1; RAIG1; feeding; fasting; GPCR;
XX KW receptor; G-protein coupled receptor; anorectic; antidiabetic;
XX KW antidepressant; immunomodulator; transgenic; gene therapy; gene; ss.
XX OS Homo sapiens.
XX PX Key Location/Qualifiers
XX FT CDS 100..1173
XX FT /*tag= a
XX FT /product= "GPCR-like RAIG1"
XX PN WO2003016553-A2.
XX PD 27-FEB-2003.
XX PX 20-AUG-2002; 2002WO-US026510.
XX PR 20-AUG-2001; 2001US-0313940P.
XX PX (GETH ) GENENTECH INC.
XX PX (CURA-) CURAGEN CORP.
XX PI Lewin DA, Stewart TA;
XX DR WPI; 2003-278580/27.
XX DR P-PSDB; ABR42649.
```

PT New G-protein coupled receptor-like retinoic acid induced gene 1 (GPCR-  
PT like RAIG1) polypeptide and gene, useful for diagnosing or treating  
PT metabolic disorders, e.g. obesity, anorexia, cachexia or diabetes.  
XX Claim 10; Page 16-17; 150pp; English.

CC The present sequence is that of human G-protein coupled receptor-like  
CC retinoic acid induced gene 1 (GPCR-like RAIG1). This is the human  
CC homologue of murine GPCR-like RAIG1. The murine gene was shown to be  
CC differentially regulated during fasting-feeding cycles in mice, with  
CC moderate induction early in fasting, down-regulation with extended  
CC fasting and 4-fold up-regulation with feeding in recovery from fasting.  
CC The differentially expressed gene, its mRNA, and the encoded protein, can  
CC each be manipulated to detect and treat metabolic disorders associated  
CC with up- or down-regulation of GPCR-like RAIG1 activity, such as obesity,  
CC anorexia, cachexia or diabetes

XX Sequence 2302 BP; 494 A; 666 C; 533 G; 609 T; 0 U; 0 Other;  
SQ Query Match 100.0%; Score 17; DB 8; Length 2302;  
Best Local Similarity 100.0%; Pred. No. 66;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCCAGCTTGCCGAG 17  
Db 1117 GCCCAGCTTGCCGAG 1133

RESULT 23  
ADD93240  
ID ADD93240 standard; cDNA; 2302 BP.

XX AC ADD93240;  
XX 29-JAN-2004 (first entry)  
XX RAIG1 coding sequence.

XX ss; Gene: retinoic acid-inducible gene 1; RAIG1; orphan;  
KW G-protein coupled receptor; GPCR; chromosome 12; GPCR5B; carcinoma;  
KW vaccine; breast cancer; pancreatic cancer; lung cancer; liver cancer;  
KW ovarian cancer; colon cancer; osteosarcoma.

XX Homo sapiens.  
XX Key Location/Qualifiers  
FT CDS 100..1173  
FT /\*tag= a  
FT /product= "RAIG1"

XX WO2003087832-A2.  
XX 23-OCT-2003.  
XX 10-APR-2003; 2003WO-GB001587.  
XX 11-APR-2002; 2002GB-00008331.  
XX 17-SEP-2002; 2002GB-00021538.  
XX (OXFO-) OXFORD GLYSCSCIENCES UK LTD.  
XX Terrett JA;  
XX WPI; 2003-845382/78.  
XX P-PSDB; ADD93239.  
XX GENBANK; AF095448.

XX Screening, diagnosing and/or treating carcinoma, including breast,  
PT pancreatic, lung, liver, ovarian and colon cancer by detecting the change  
PT in expression or activity of an RAIG1 polypeptide or encoding nucleic  
PT acid molecule.  
XX Claim 1; Fig 2; 43pp; English.

XX This sequence encodes retinoic acid-inducible gene 1 (RAIG1) polypeptide.  
CC RAIG1 is an orphan G-protein coupled receptor (GPCR) located on  
CC chromosome 12. RAIG1 shows a restricted expression pattern compared to a  
CC related receptor, GPCR5B, which is widely expressed in peripheral and  
CC central tissues. The RAIG1 cDNA and polypeptide sequence may be used in  
CC the method of the invention for screening for and/or diagnosis of  
CC carcinoma in a subject, and/or monitoring the effectiveness of carcinoma  
CC therapy. The method comprises detecting and/or quantifying in a  
CC biological sample obtained from the subject an RAIG1 polypeptide and a  
CC nucleic acid molecule. The RAIG1 polypeptide and nucleic acid molecule  
CC are useful in the manufacture of a medicament for the treatment of  
CC carcinoma, where the composition is a vaccine. An agent which interacts  
CC with or causes change in the expression or activity of an RAIG1  
CC polypeptide or nucleic acid molecule, is also useful in the manufacture  
CC of a medicament for the treatment of carcinoma that is breast cancer,  
CC pancreatic cancer, lung cancer, liver cancer, ovarian cancer, colon  
CC cancer and/or osteosarcoma. They can also be used in the diagnosis and  
CC screening of such carcinomas.

XX Sequence 2302 BP; 494 A; 666 C; 533 G; 609 T; 0 U; 0 Other;  
SQ Query Match 100.0%; Score 17; DB 10; Length 2302;  
Best Local Similarity 100.0%; Pred. No. 66;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCCAGCTTGCCGAG 17  
Db 1117 GCCCAGCTTGCCGAG 1133

RESULT 24  
ADL24773  
ID ADL24773 standard; DNA; 2302 BP.

XX AC ADL24773;  
XX 20-MAY-2004 (first entry)  
XX Intestinal epithelium/peyer's patch M cell-associated DNA sequence #171.

XX Intestinal epithelium cell development; peyer's patch M cell development;  
KW inflammatory bowel disease; glutenenteropathy; infectious disease;  
KW autoimmune disease; haemolytic anaemia; rheumatoid arthritis; dermatitis;  
KW Grave's disease; multiple sclerosis; allergy; asthma; diabetic mellitus;  
KW immune system disorder; hypersensitivity; anaphylaxis;  
KW blood group incompatibility; ds; human.

XX Homo sapiens.  
XX WO200280852-A2.  
XX 17-OCT-2002.  
XX 04-APR-2002; 2002WO-US010873.  
XX 04-APR-2001; 2001US-0281416P.  
XX (DIGI-) DIGITAL GENE TECHNOLOGIES INC.  
XX Brayden DJ, Byrne D, O'mahony DJ, Evans CF, Mah SP, Lo DD;  
XX WPI; 2003-075470/07.  
XX Novel isolated or purified polypeptide encoded by genes associated with  
PT intestinal epithelium or M cell development, differentiation or function,  
PT useful for treating autoimmune diseases and infectious diseases.  
XX Claim 1; SEQ ID NO 283; 152pp; English.

XX The invention comprises DNA sequences which are associated with  
CC intestinal epithelium and peyer's patch M cells. The DNA sequences of the  
CC invention are useful for assessing, modifying, modulating or regulating





RESULT 27  
ABZ42832  
ID ABZ42832 standard; DNA; 2456 BP.  
XX AC ABZ42832;  
XX DT 04-MAR-2003 (first entry)  
XX DE Human G protein-coupled receptor RAIG1 nucleotide SEQ ID NO:453.  
XX KW G protein-coupled receptor; GPCR; antigenic peptide; gene therapy;  
KW G protein-coupled receptor modulator; antibody; immune-related disease;  
KW growth-related disease; cell regeneration-related disease; AIDS; cancer;  
KW immunological-related cell proliferative disease; autoimmune disease;  
KW Alzheimer's disease; atherosclerosis; infection; osteoarthritis; allergy;  
KW osteoporosis; cardiomyopathy; inflammation; Crohn's disease; diabetes;  
KW graft versus host disease; Parkinson's disease; multiple sclerosis; pain;  
KW psoriasis; anxiety; depression; schizophrenia; dementia; memory loss;  
KW mental retardation; epilepsy; asthma; tuberculosis; obesity; nausea;  
KW hypertension; hypotension; renal disorder; rheumatoid arthritis; trauma;  
ulcer; gene; ds.  
XX OS Homo sapiens.  
XX PN WO200261087-A2.  
XX PD 08-AUG-2002.  
XX PF 19-DEC-2001; 2001WO-US050107.  
XX PR 19-DEC-2000; 2000US-0257144P.  
XX PA (LIFE-) LIFESPAN BIOSCIENCES INC.  
XX PI Burner GC, Roush CL, Brown JP;  
XX WPI; 2003-046718/04.  
XX P-PSDB; ABP81984.  
XX New isolated antigenic peptides e.g., for G protein-coupled receptors  
(GPCR), useful for diagnosing and designing drugs for treating conditions  
in which GPCRs are involved, e.g. AIDS, Alzheimer's disease, cancer or  
autoimmune diseases.  
XX Disclosure; Fig 1; 523pp; English.  
XX The present invention describes antigenic peptides (I) comprising: (a)  
any one of 1601 sequences (see ABP82019 to ABP83619) of 12-24 amino  
acids. Also described: (1) an assay for the detection of a particular G  
protein-coupled receptor (GPCR) or a candidate polypeptide in a sample;  
and (2) an isolated antibody having high specificity and high affinity or  
avidity for a particular GPCR. (I) can be used as GPCR modulators and in  
gene therapy. The antigenic peptides for GPCRs are useful in detecting an  
antibody against a particular GPCR, and in the production of specific  
antibodies. The peptides and antibodies are also useful for detecting the  
presence or absence of corresponding GPCRs. The antigenic peptides for  
GPCRs and antibodies are useful for diagnosing and designing drugs for  
treating immune-related diseases, growth-related diseases, cell  
regeneration-related disease, immunological-related cell proliferative  
diseases, or autoimmune diseases, e.g. AIDS, Alzheimer's disease,  
atherosclerosis, bacterial, fungal, protozoan or viral infections,  
osteoarthritis, osteoporosis, cancer, cardiomyopathy, chronic and acute  
inflammation, allergies, Crohn's disease, diabetes, graft versus host  
disease, Parkinson's disease, multiple sclerosis, pain, psoriasis,  
anxiety, depression, schizophrenia, dementia, mental retardation, memory  
loss, epilepsy, asthma, tuberculosis, obesity, nausea, hypertension,  
hypotension, renal disorders, rheumatoid arthritis, trauma, ulcers, or  
any other disorder in which GPCRs are involved. The antibodies may be  
used in immunoassays and immunodiagnosis. ABZ42523 to ABZ42869 encode  
GPCR proteins given in ABP81675 to ABP82018, which are used in the  
exemplification of the present invention

XX SQ Sequence 2456 BP; 519 A; 722 C; 571 G; 644 T; 0 U; 0 Other;  
Query Match 100.0%; Score 17; DB 8; Length 2456;  
Best Local Similarity 100.0%; Pred. No. 66;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GCCACGCTTGCCGAG 17  
DB 1271 GCCACGCTTGCCGAG 1287  
RESULT 28  
ACC72695  
ID ACC72695 standard; cDNA; 2456 BP.  
XX AC ACC72695;  
XX DT 09-JUL-2003 (first entry)  
XX DE Human cancer related protein encoding cDNA SEQ ID NO:34.  
XX KW Human; cancer; diagnosis; screening; modulator; leukaemia; ischaemia;  
KW heart disease; atherosclerosis; endometriosis; gene; ss.  
XX OS Homo sapiens.  
XX PN WO2003025138-A2.  
XX PD 27-MAR-2003.  
XX PF 17-SEP-2002; 2002WO-US029560.  
XX PR 17-SEP-2001; 2001US-0323469P.  
PR 20-SEP-2001; 2001US-0323887P.  
PR 13-NOV-2001; 2001US-0350666P.  
PR 08-FEB-2002; 2002US-0355145P.  
PR 08-FEB-2002; 2002US-0355257P.  
PR 12-APR-2002; 2002US-0372246P.  
XX PA (EOSB-) EOS BIOTECHNOLOGY INC.  
XX AFar D, Aziz N, Gish KC, Hevezi PA, Mack DH, Wilson KE;  
PI Zlotnik A;  
XX WPI; 2003-354600/33.  
XX P-PSDB; ABR58568.  
XX New genes that are up-regulated or down-regulated in cancers, useful as  
markers for diagnosing e.g. cancer, ischemia or heart diseases, or as  
therapeutic targets for screening drugs for treating these diseases.  
XX Claim 8; Page 643-644; 767pp; English.  
XX The present invention describes an isolated nucleic acid molecule, which  
comprises the sequence of any of the genes that are up-regulated or down-  
regulated in specific cancers (e.g. about 1031 genes up-regulated in  
acute lymphocytic leukemia). ACC72641 to ACC72860 represent cancer  
related gene nucleotide sequences which encode the proteins given in  
ABR58521 to ABR58709. Also described: (1) determining the presence or  
absence of a pathological cell in a patient; (2) an expression vector  
comprising a nucleic acid molecule described above; (3) a host cell  
comprising the vector; (4) an isolated polypeptide, which is encoded by  
the nucleic acid; (5) an antibody that specifically binds the polypeptide  
of (4); (6) specifically targeting a compound to a pathological cell in a  
patient by administering to the patient the antibody above; and (7) a  
drug screening assay. The nucleic acid is useful as diagnostic markers or  
therapeutic targets. In particular, the nucleic acid is useful for  
diagnosing a pathology, e.g. cancer (e.g. cancer of the bone marrow,  
bladder, brain, breast, cervix, colon/rectum, kidney, lung, ovary,  
pancreas, prostate, skin and uterus), wounds, ischaemia, heart diseases,  
atherosclerosis and endometriosis. The nucleic acid is also useful in  
drug screening, particularly for identifying agents for treating these

CC pathologies  
XX Sequence 2456 BP; 522 A; 720 C; 571 G; 643 T; 0 U; 0 Other;  
SQ Query Match 100.0%; Score 17; DB 10; Length 2456;  
Best Local Similarity 100.0%; Pred. No. 66;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17  
Db 1271 GCCCAGCGTTGGCCGAG 1287

RESULT 29  
ADN39301  
ID ADN39301 standard; cDNA; 2456 BP.  
XX  
AC ADN39301;  
XX  
DT 17-JUN-2004 (first entry)  
XX  
DE Cancer/angiogenesis/fibrosis-related nucleic acid, SEQ ID NO:619.  
XX  
KW Human; differential expression; cancer; angiogenic disorder;  
KW fibrotic disorder; psoriasis; ischaemia; heart disease; atherosclerosis;  
KW inflammatory disease; autoimmune disease;  
KW retinal neovascularisation syndrome; scarring; uterine fibroid;  
KW detection; diagnosis; prognosis; drug screening; drug targeting;  
KW wound healing; contraception; cytostatic; cardiac; immunomodulatory;  
KW vulnerable; gene therapy; vaccine; gene; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO2003042661-A2.  
XX  
PD 22-MAY-2003.  
XX  
XX 13-NOV-2002; 2002WO-US036810.  
XX  
XX 13-NOV-2001; 2001US-0350666P.  
XX  
XX 21-NOV-2001; 2001US-0332464P.  
XX  
XX 29-NOV-2001; 2001US-0334393P.  
XX  
XX 03-DEC-2001; 2001US-0335394P.  
XX  
XX 14-DEC-2001; 2001US-0340376P.  
XX  
XX 08-JAN-2002; 2002US-0347211P.  
XX  
XX 10-JAN-2002; 2002US-0347349P.  
XX  
XX 08-FEB-2002; 2002US-035250P.  
XX  
XX 13-FEB-2002; 2002US-0356714P.  
XX  
XX 20-FEB-2002; 2002US-0359077P.  
XX  
XX 29-MAR-2002; 2002US-0368809P.  
XX  
XX 04-APR-2002; 2002US-0370110P.  
XX  
XX 12-APR-2002; 2002US-0372246P.  
XX  
XX 05-JUN-2002; 2002US-0386614P.  
XX  
XX 16-JUL-2002; 2002US-0396839P.  
XX  
XX 22-JUL-2002; 2002US-039775P.  
XX  
XX 22-JUL-2002; 2002US-0397845P.  
XX  
XX 09-SEP-2002; 2002US-0409450P.  
XX  
XX (EOSB-) EOS BIOTECHNOLOGY INC.  
XX  
XX Afar D, Aziz N, Ginsburg WM, Gish KC, Glynn R, Hevezi PA;  
PI Mack DH, Murray R, Watson SR, Wilson KE, Zlotnik A;  
XX  
XX WPI; 2003-468649/44.  
XX  
XX P-PSDB; ADN39302.  
XX  
XX  
XX Determining the presence or absence of a pathological cell in a patient,  
FT useful for diagnosing, prognosing or treating cancer, comprises detecting  
FT a nucleic acid in a biological sample.  
XX  
XX Claim 8; SEQ ID NO 619; 1385pp; English.  
XX  
XX The invention relates to nucleic acids and proteins (ADN38683-ADN40064)

CC whose expression is upregulated or downregulated in specific cancers or  
CC other diseases such as angiogenic or fibrotic disorders, and to methods  
CC of determining the presence or absence of a pathological cell in a  
CC patient by detecting a nucleic acid at least 80% identical to those of  
CC the invention or by detecting a polypeptide of the invention. The  
CC invention also relates to expression vectors and host cells comprising a  
CC nucleic acid of the invention; antibodies which specifically bind a  
CC polypeptide of the invention; use of such antibodies for drug targeting;  
CC and methods of screening for modulators of activity or expression of the  
CC polypeptides and nucleic acids. The nucleic acids, polypeptides,  
CC antibodies and methods are useful for diagnosing, prognosing and treating  
CC cancer and other conditions such as psoriasis, ischaemia, heart disease,  
CC atherosclerosis, inflammatory diseases, autoimmune diseases, retinal  
CC neovascularisation syndromes, scarring and uterine fibroids. They may  
CC also be useful in wound healing and in contraception. The present  
CC sequence represents a nucleic acid sequence of the invention.  
XX  
SQ Sequence 2456 BP; 519 A; 722 C; 571 G; 644 T; 0 U; 0 Other;  
Query Match 100.0%; Score 17; DB 11; Length 2456;  
Best Local Similarity 100.0%; Pred. No. 66;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17  
Db 1271 GCCCAGCGTTGGCCGAG 1287

RESULT 30  
ADI28525  
ID ADI28525 standard; cDNA; 2456 BP.  
XX  
AC ADI28525;  
XX  
DT 22-APR-2004 (first entry)  
XX  
DE Human GPCR retinoic acid induced 3 (RAI-3) cDNA.  
XX  
XX Retinoic acid induced 3; RAI-3; human; G-protein coupled receptor; GPCR;  
KW antiinflammatory; immunosuppressive; cytostatic; cardiac; antiallergic;  
KW broncholytic; gene therapy; gene; single nucleotide polymorphism; SNP;  
KW chromosome 12p13-p12.3; ss.  
XX  
OS Homo sapiens.  
XX  
XX Key Location/Qualifiers  
FT variation replace(112,r)  
FT /tag= a  
FT /label= RAI-3-s1  
FT /note= "located in 5' untranslated region"  
FT /standard\_name= "single nucleotide polymorphism"  
FT 254..1327  
FT /tag= b  
FT /product= "Human RAI3"  
FT variation replace(364,y)  
FT /tag= c  
FT /label= RAI-3-s2  
FT /note= "exon 1, silent (Ala/Ala)"  
FT /standard\_name= "single nucleotide polymorphism"  
FT variation replace(511,y)  
FT /tag= d  
FT /label= RAI-3-s3  
FT /note= "exon 2, silent (Ile/Ile)"  
FT /standard\_name= "single nucleotide polymorphism"  
FT variation replace(523,y)  
FT /tag= e  
FT /label= RAI-3-s4  
FT /note= "exon 2, silent (Asp/Asp)"  
FT /standard\_name= "single nucleotide polymorphism"  
FT variation replace(605,r)  
FT /tag= f  
FT /label= RAI-3-s6  
FT /note= "exon 2, missense (Ser/Gly)"  
FT

```
FT /standard_name= "Single nucleotide polymorphism"
FT replace(797,r)
FT /*tag= g
FT /label= RAI-3-s5
FT /note= "exon 2, missense (Thr/Ala)"
FT /standard_name= "Single nucleotide polymorphism"
FT replace(1111,t)
FT /*tag= h
FT /label= RAI-3-s8
FT /note= "silent (Pro/Pro)"
FT /standard_name= "Single nucleotide polymorphism"
FT replace(1173,t)
FT /*tag= i
FT /label= RAI-3-s9
FT /note= "missense (Gln/Arg)"
FT /standard_name= "Single nucleotide polymorphism"
XX WO2004001060-A2.
PN
XX
XX 31-DEC-2003.
PD
XX
XX 20-JUN-2003; 2003WO-US019255.
PP
XX
XX 20-JUN-2002; 2002US-0390850P.
PR
XX 29-AUG-2002; 2002US-0407006P.
PR
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.
PA
XX
XX Whitney GS, Opitreck G, Garulacan L, Ramanathan CS, McKinnon M;
PI Bennett KL, Barber LE, Cacace A, Tsuchihashi Z;
XX
XX WPI; 2004-090973/09.
DR
XX P-PSDB; ADI28526.
DR
XX
XX New nucleic acid molecule encoding a human G-protein coupled
FT receptor, RAI3, useful for preventing, treating or ameliorating chronic obstructive
FT pulmonary disease (COPD), COPD-like disorder, or the underlying symptoms
FT of COPD.
XX
XX Claim 21; SEQ ID NO 18; 301pp; English.
PS
XX
XX The present sequence is that of cDNA encoding a human G-protein coupled
CC receptor (Class C, Group 5), denoted retinoic acid induced 3 (RAI-3).
CC Proteomics methods were used to isolate cigarette smoke-inducible
CC tyrosine phosphorylated proteins from airway epithelial cells. RAI-3 was
CC identified as being tyrosine phosphorylated and/or as being
CC associated/complexed with tyrosine phosphorylated proteins only in those
CC cells that had been exposed to cigarette smoke. Since RAI-3 is primarily
CC expressed in lung tissue, and since cigarette smoke is a major causative
CC factor of chronic obstructive pulmonary disease (COPD), RAI-3 provides a
CC novel cellular target for identifying modulators, e.g. agonists or
CC antagonists, useful for the treatment and/or prevention of COPD and
CC related disorders such as emphysema and chronic bronchitis. RAI-3
CC modulators, e.g. agonists and antagonists, especially antisense
CC compounds, can be used to treat COPD and/or its associated and diseases
CC signaling molecules. Single nucleotide polymorphisms (SNPs) detected in
CC the RAI-3 gene are useful for determining COPD association in
CC individuals. RAI-3 nucleic acid molecules and polypeptides are useful for
CC preventing, treating or ameliorating disorders related to aberrant GPCR
CC signaling or cell cycle regulation, pulmonary disorders, inflammatory
CC lung disorders, COPD, the underlying symptoms of COPD, COPD-related
CC disorders or conditions, autoimmune disorders, disorders related to
CC hyperimmune activity, inflammatory conditions, disorders related to
CC aberrant acute phase responses, hypercongenital conditions, birth
CC defects, necrotic lesions, wounds, organ transplant rejection, renal
CC diseases, ischaemia-reperfusion injury, heart disorders, disorders
CC related to aberrant signal transduction, proliferation disorders,
CC cancers, HIV infection, asthma, cystic fibrosis, pulmonary fibrosis,
CC ulcerative colitis, cerebral infarct, myocardial infarct, diabetic
CC nephropathy, allergic rhinitis, Crohn's disease, atherosclerosis,
CC rheumatoid arthritis, inflammatory/autoimmune disorders, glioblastoma,
CC pulmonary small cell undifferentiated carcinoma, carcinoma of the breast,
```

```
CC colon, lung, ovary, pancreas, prostate, non-Hodgkin's lymphoma, disorders
CC associated with aberrant cell adhesion, I-CAM function and/or regulation,
CC E-selectin function and/or regulation, or aberrant NF-kB function and/or
CC regulation (all claimed).
XX
SQ Sequence 2456 BP; 516 A; 719 C; 570 G; 643 T; 0 U; 8 Other;
Query Match 100.0%; Score 17; DB 12; Length 2456;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCCCAGCTTGGCCGAG 17
Db 1271 GCCCAGCTTGGCCGAG 1287
RESULT 31
ADI28459
ID ADI28459 standard; cDNA; 2456 BP.
XX
XX ADI28459;
XX
XX 22-APR-2004 (first entry)
DT
XX Human GPCR retinoic acid induced 3 (RAI-3) cDNA.
DE
XX
XX Retinoic acid induced 3; RAI-3; human; G-protein coupled receptor; GPCR;
KW antiinflammatory; immunosuppressive; cytosolic; cardiac; anti-allergic;
KW broncholytic; gene therapy; gene; single nucleotide polymorphism; SNP;
KW chromosome 12p13-p12.3; ss.
XX
XX Homo sapiens.
XX
XX FH Key Location/Qualifiers
FT variation replace(112,a)
FT /*tag= a
FT /label= RAI-3-s1
FT /note= "located in 5' untranslated region"
FT /standard_name= "Single nucleotide polymorphism"
FT CDS 254..1327
FT /*tag= b
FT /product= "Human RAI3"
FT variation replace(364,t)
FT /*tag= c
FT /label= RAI-3-s2
FT /note= "exon 1, silent (Ala/Ala)"
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(511,t)
FT /*tag= d
FT /label= RAI-3-s3
FT /note= "exon 2, silent (Ile/Ile)"
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(523,t)
FT /*tag= e
FT /label= RAI-3-s4
FT /note= "exon 2, silent (Asp/Asp)"
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(605,g)
FT /*tag= f
FT /label= RAI-3-s6
FT /note= "exon 2, missense (Ser/Gly)"
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(797,g)
FT /*tag= g
FT /label= RAI-3-s5
FT /note= "exon 2, missense (Thr/Ala)"
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(1111,g)
FT /*tag= h
FT /label= RAI-3-s8
FT /note= "silent (Pro/Pro)"
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(1173,g)
```

```
FT FT /*tag= i
FT FT /label= RAI-3-69
FT FT /note= "missense (Gln/Arg)"
FT FT /standard_name= "Single nucleotide polymorphism"
XX PN WO2004001060-A2.
XX XX 31-DEC-2003.
XX XX 20-JUN-2003; 2003WO-US019255.
XX XX 20-JUN-2002; 2002US-0390850P.
XX XX 29-AUG-2002; 2002US-0407006P.
XX XX (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX XX Whitney GS, Opitck G, Garulacan L, Ramanathan CS, McKinnon M;
XX XX Bennett KL, Barber LE, Cacace A, Tsuchihashi Z;
XX XX WPI; 2004-090973/09.
XX XX P-PSDB; ADI28460.
XX XX GENBANK; NM_003979.
XX XX New nucleic acid molecule encoding a human G-protein coupled receptor,
XX XX RAI3, useful for preventing, treating or ameliorating chronic obstructive
XX XX pulmonary disease (COPD), COPD-like disorder, or the underlying symptoms
XX XX of COPD.
XX XX Claim 1; SEQ ID NO 2; 301pp; English.
XX XX The present sequence is that of cDNA encoding a human G-protein coupled
XX XX receptor (Class C, Group 5), denoted retinoic acid induced 3 (RAI-3).
XX XX Proteomic methods were used to isolate cigarette smoke-inducible
XX XX tyrosine phosphorylated proteins from airway epithelial cells. RAI-3 was
XX XX identified as being tyrosine phosphorylated and/or as being
XX XX associated/complexed with tyrosine phosphorylated proteins only in those
XX XX cells that had been exposed to cigarette smoke. Since RAI-3 is primarily
XX XX expressed in lung tissue, and since cigarette smoke is a major causative
XX XX factor of chronic obstructive pulmonary disease (COPD), RAI-3 provides a
XX XX novel cellular target for identifying modulators, e.g. agonists or
XX XX antagonists, useful for the treatment and/or prevention of COPD and
XX XX related disorders such as emphysema and chronic bronchitis. RAI-3
XX XX modulators, e.g. agonists and antagonists, especially antisense
XX XX compounds, can be used to treat COPD and other disorders and diseases
XX XX associated with regulation of NF-kB and/or its associated or interacting
XX XX signaling molecules. Single nucleotide polymorphisms (SNPs) detected in
XX XX the RAI-3 gene are useful for determining COPD association in
XX XX individuals. RAI-3 nucleic acid molecules and polypeptides are useful for
XX XX preventing, treating or ameliorating disorders related to aberrant GPCR
XX XX signaling or cell cycle regulation, pulmonary disorders, inflammatory
XX XX lung disorders, COPD, the underlying symptoms of COPD, COPD-related
XX XX disorders or conditions, autoimmune disorders, disorders related to
XX XX hyperimmune activity, inflammatory conditions, disorders related to
XX XX aberrant acute phase responses, hypercongenital conditions, birth
XX XX defects, necrotic lesions, wounds, organ transplant rejection, renal
XX XX diseases, ischemia-reperfusion injury, heart disorders, disorders
XX XX related to aberrant signal transduction, proliferation disorders,
XX XX cancers, HIV infection, asthma, cystic fibrosis, pulmonary fibrosis,
XX XX ulcerative colitis, cerebral infarct, myocardial infarct, diabetic
XX XX nephropathy, allergic rhinitis, Crohn's disease, atherosclerosis,
XX XX rheumatoid arthritis, inflammatory/autoimmune disorders, glioblastoma,
XX XX pulmonary small cell undifferentiated carcinoma, carcinoma of the breast,
XX XX colon, lung, ovary, pancreas, prostate, non-Hodgkin's lymphoma, disorders
XX XX associated with aberrant cell adhesion, I-CAM function and/or regulation,
XX XX E-selectin function and/or regulation, or aberrant NF-kB function and/or
XX XX regulation (all claimed).
XX XX
XX XX Query Match 100.0%; Score 17; DB 12; Length 2456;
XX XX Best Local Similarity 100.0%; Pred. No. 66;
XX XX Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
SQ Sequence 2456 BP; 519 A; 722 C; 571 G; 644 T; 0 U; 0 Other;
Query Match 100.0%; Score 17; DB 13; Length 2456;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GCCCAGCGTTGCCGCGAG 17
|||||
Db 1271 GCCCAGCGTTGCCGCGAG 1287

RESULT 33
ACN39388
ID ACN39388 standard; cDNA; 2456 BP.
XX
AC ACN39388;
XX
DT 18-NOV-2004 (first entry)
XX
DE Tumour-associated antigenic target (TAT) cDNA DNA226771, SEQ ID NO:3498.
XX
KW Tumour-associated antigenic target; TAT; human; overexpression; cancer;
KW tumour; diagnosis; cell proliferative disorder; breast cancer;
KW colorectal cancer; lung cancer; ovarian cancer; liver cancer;
KW central nervous system cancer; bladder cancer; pancreatic cancer;
KW cervical cancer; melanoma; leukaemia; hybridisation probe;
KW chromosome identification; chromosome mapping; gene mapping;
KW gene therapy; cytostatic; gene; ss.
XX
OS Homo sapiens.
XX
PN WO2004030615-A2.
XX
PD 15-APR-2004.
XX
PF 29-SEP-2003; 2003WO-US028547.
XX
PR 02-OCT-2002; 2002US-0414971P.
XX
PA (GETH ) GENENTECH INC.
XX
PI Wu TD, Zhang Z, Zhou Y;
XX
WPI: 2004-347921/32.
XX
P-PSDB; ABM81354.
XX
New tumor-associated antigenic target polypeptides and nucleic acids,
PT useful in preparing a medicament for treating or detecting a
PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or
PT prostate cancer or tumor.
XX
PS Claim 1; SEQ ID NO 3498; 7273pp; English.
XX
The invention relates to human tumour-associated antigenic target (TAT)
CC polypeptides, and their related nucleic acids. The TAT polypeptides are
CC overexpressed in cancer tissues compared to normal tissues, and may thus
CC serve as effective targets for the diagnosis and treatment of cancer in
CC mammals. The invention also relates to nucleic acid and polypeptide
CC sequences at least 80% identical to the TAT nucleic acids and
CC polypeptides; expression vectors and host cells comprising a TAT nucleic
CC acid; an antibody specific for a TAT polypeptide; a peptide or organic
CC molecule which binds to a TAT polypeptide; fusion proteins comprising a
CC TAT polypeptide; and methods and compositions for the treatment or
CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
CC antibodies, antagonists, binding molecules and compositions are useful
CC for diagnosing or treating a cell proliferative disorder associated with
CC increased TAT expression, particularly cancers such as breast cancer,
CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
CC cancer, pancreatic cancer, cervical cancer, cancers of the central
CC nervous system, melanoma and leukaemia. TAT nucleic acids may further be
CC used as hybridisation probes, in chromosome and gene mapping, in
CC chromosome identification and in gene therapy. The present sequence
CC represents a TAT nucleic acid of the invention
```

---

```
XX
SQ Sequence 2456 BP; 519 A; 722 C; 571 G; 644 T; 0 U; 0 Other;
Query Match 100.0%; Score 17; DB 13; Length 2456;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GCCCAGCGTTGCCGCGAG 17
|||||
Db 1271 GCCCAGCGTTGCCGCGAG 1287

RESULT 34
ADR43992
ID ADR43992 standard; DNA; 2456 BP.
XX
AC ADR43992;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human breast tumour associated gene clone-02 SEQ ID NO:2.
XX
KW cancer; tumour; breast cancer; lung cancer; colon cancer; kidney cancer;
KW antibody; cytostatic; gene therapy; human; breast tumour tissue;
KW breast tissue; gene; ds.
XX
OS Homo sapiens.
XX
PN WO2004074506-A2.
XX
PD 02-SEP-2004.
XX
PF 30-JAN-2004; 2004WO-US003030.
XX
PR 13-FEB-2003; 2003US-0447900P.
XX
PA (MERG-) MERGEN LTD.
XX
PI Hu Q, Peng A, Liu B, Love JR, Hao X, Ren M, Sheng Z;
XX
WPI; 2004-635589/61.
XX
Diagnosing and treating a cancer, e.g. breast, lung, colon, or kidney
PT cancer, comprises detecting and blocking the over expression of a gene of
PT a protein found in breast, lung, colon, or kidney tissue.
XX
PS Claim 2; SEQ ID NO 2; 220pp; English.
XX
The present invention describes a method for diagnosing and treating a
CC cancer, e.g. breast, lung, colon, or kidney cancer. The method comprises
CC detecting and blocking the over expression of a gene of a protein found
CC in breast, lung, colon, or kidney tissue. Also described is an antibody
CC or binding portion of an antibody that specifically binds a protein found
CC in breast tissue, lung tissue, colon tissue, or kidney tissue. The
CC antibody has cytostatic activity, and can be used in gene therapy. The
CC methods, antibodies, polynucleotides and polypeptides from the present
CC invention are useful for detecting, diagnosing, preventing and treating
CC cancer, e.g. breast, lung, colon, or kidney cancer. The present sequence
CC represents a nucleotide sequence given in the present invention, which is
CC over expressed in human breast tumour tissue. N.B. All 385 sequences
CC referenced in this patent are detailed in the US provisional application
CC SN 60/447,900, filed 02/13/2003. In this application only one sequence,
CC representing the longest sequence of each of the 65 clones is listed in
CC Appendix A (and given as SEQ ID NO:1 to 65 in the Sequence Listing).
XX
SQ Sequence 2456 BP; 517 A; 710 C; 567 G; 643 T; 0 U; 19 Other;
Query Match 100.0%; Score 17; DB 13; Length 2456;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GCCCAGCGTTGCCGCGAG 17
|||||
```

```
Db      1271 GCCACGCTTGCCGAG 1287
RESULT 35
ADU06126
ID      ADU06126 standard; DNA; 2456 BP.
XX
AC      ADU06126;
XX
XX      27-JAN-2005 (first entry)
DT
XX      Novel bronchial cancer-associated human gene SeqID350.
DE
XX      bronchial cancer; cytostatic; tumour-associated protein;
KW      cancer detection; metastasis; tumour; gene; ds; human.
XX
XX      Homo sapiens.
OS
XX      DE10316701-A1.
PN
XX      04-NOV-2004.
PD
XX      09-APR-2003; 2003DE-01016701.
PF
XX      09-APR-2003; 2003DE-01016701.
PR
XX      (HINZ/) HINZMANN B.
PA      (HERM/) HERMANN K.
PA      (CAST/) HEIDEN CASTANOS-VELEZ E.
XX
XX      Mennerich D, Bruemendorf T, Heiden E, Hermann K, Kinnenmann H;
PI      Li X, Roepcke S, Staub E, Hinzmann B, Rosenthal A, Pflarsky C;
XX
XX      WPI; 2004-786403/78.
DR      P-PSDB; ADU06613.
XX
XX      New nucleic acid, and derived proteins, useful for diagnosis of bronchial
PT      cancer and in screening for therapeutic and diagnostic agents.
XX
XX      Claim 1; SEQ ID NO 350; 1381pp; German.
PS
XX
XX      This invention relates to a novel isolated nucleic acid associated with
CC      bronchial cancer comprising 489 defined sequences given in the
CC      specification. The invention may be useful for the production of
CC      compounds with a cytostatic activity through the inhibition of expression
CC      or activity of tumour-associated proteins. The novel DNA sequences and
CC      the proteins/peptides encoded by them are used for detecting bronchial
CC      cancer or determining the risk of developing it and to screen for
CC      specific binding partners of the DNA or protein sequences, where the
CC      binding partners are potentially useful as agents for treating or
CC      diagnosing bronchial cancer. The DNA or protein sequences can also be
CC      used for prognosis, detection of metastases and for secondary treatment
CC      (of tumours that have been stabilised or are no longer detectable).
CC      Detecting abnormal expression of the DNA sequences provides early
CC      diagnosis of bronchial cancers. The present sequence is that of a novel
CC      bronchial cancer-associated human gene sequence of the invention.
XX
XX      Sequence 2456 BP; 519 A; 722 C; 571 G; 644 T; 0 U; 0 Other;
SQ
Query Match      100.0%; Score 17; DB 13; Length 2456;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy      1 GCCACGCTTGCCGAG 17
      |||||
Db      1271 GCCACGCTTGCCGAG 1287
RESULT 36
AEA00088
ID      AEA00088 standard; cDNA; 2456 BP.
XX
AC      AEA00088;
XX
XX      28-JUL-2005 (first entry)
DT
XX      Human TAT115 cDNA SEQ ID NO:40.
DE
XX      ss; gene; tumor-associated antigen; cytostatic; breast tumor;
KW      endocrine disease; gynecology and obstetrics; neoplasm; colon tumor;
KW      gastrointestinal disease; rectal tumor; endometroid carcinoma;
KW      genitourinary disease; renal tumor; lung tumor; respiratory disease;
KW      ovary tumor; skin tumor; liver tumor.
XX
XX      Homo sapiens.
OS
XX      Key Location/Qualifiers
PH      254..1327
FT      /*tag= a
FT      /product= "TAT115"
XX
XX      US2005106644-A1.
PN
XX      19-MAY-2005.
PD
XX      08-SEP-2004; 2004US-00936626.
PF
XX      20-JUN-2001; 2001US-0299500P.
PR      29-JUN-2001; 2001US-0301880P.
PR      18-SEP-2001; 2001US-0323268P.
PR      19-JUN-2002; 2002US-00177488.
PR      26-MAR-2004; 2004US-0557116P.
PR      04-AUG-2004; 2004US-0598899P.
XX
XX      (SETH ) GENENTECH INC.
PA
XX      Cairns B, Chen R, Frantz G, Hillan KJ, Koepfen H, Phillips HS;
PI      Polakis P, Spencer SD, Smith V, Williams PM, Wu TD, Zhang Z;
PI      Sliwkowski M;
XX
XX      WPI; 2005-384304/39.
DR      P-PSDB; AEA00166.
XX
XX      Novel isolated antibody capable of binding to tumor-associated antigenic
PT      target polypeptide, useful for treating cell proliferative disorder e.g.
XX      cancer.
PS      Claim 1; SEQ ID NO 40; 337pp; English.
XX
XX      The invention relates to a novel isolated antibody binding to a
CC      polypeptide having at least 80% sequence identity to a polypeptide having
CC      any one of 76 fully defined 182-910 amino acid tumor-associated antigenic
CC      target polypeptide (TAT) sequences (AEA00127-AEA00202) given in the
CC      specification, a polypeptide having any one of (AEA00127-AEA00202), a
CC      lacking its associated signal peptide, or an extracellular domain of a
CC      polypeptide having any one of (AEA00127-AEA00202). The polypeptide is
CC      encoded by the nucleotide sequence having any one of (AEA00049-AEA00126).
CC      An antibody of the invention has cytostatic activity. The antibody is
CC      useful for inhibiting growth of a cell expressing TAT188, which involves
CC      contacting the cell with the antibody. The cell is a cancer cell chosen
CC      from breast, colon, rectum, endometrium, kidney, lung, ovary, skin and
CC      liver cell. The cancer cell is a mammalian cell, preferably a human cell.
CC      The antibody is also useful for detecting the level of TAT188
CC      polypeptide expressed in a test cell relative to a control cell, and for
CC      detecting the level of TAT188 polypeptide or a polypeptide having at
CC      least 80 % sequence identity to the TAT188 polypeptide sequence in a test
CC      cell relative to a control cell. The antibody is useful for treating a
CC      preventing a cell proliferative disorder associated with increased
CC      expression or activity of a polypeptide having at least 80 % identity to
CC      a TAT188 polypeptide sequence. The cell proliferative disorder is cancer.
CC      The method of the invention is useful for inhibiting the growth of a
CC      cancer cell. The present sequence encodes a polypeptide of the invention.
XX
XX      Sequence 2456 BP; 519 A; 722 C; 571 G; 644 T; 0 U; 0 Other;
SQ
Query Match      100.0%; Score 17; DB 14; Length 2456;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Best Local Similarity 100.0%; Pred. No. 66;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACGCTTGGCCGAG 17
Db 1271 GCCACGCTTGGCCGAG 1287

RESULT 37
AEA00608
ID AEA00608 standard; cDNA; 2456 BP.
XX
AC AEA00608;
XX
DT 28-JUL-2005 (first entry)
XX
DE Human TAT115 cDNA sequence SeqID40.
XX
KW antibody identification; tumor-associated antigen; cytostatic;
KW RNA interference; gene therapy; cell death; cancer; breast tumor;
KW colon tumor; rectal tumor; renal tumor; lung tumor; ovary tumor;
KW skin tumor; liver tumor; gene; ss; TAT115.
XX
OS Homo sapiens.
XX
PN US2005107595-A1.
XX
PD 19-MAY-2005.
XX
PF 10-SEP-2004; 2004US-00938061.
XX
PR 20-JUN-2001; 2001US-0299500P.
PR 29-JUN-2001; 2001US-0301880P.
PR 18-SEP-2001; 2001US-0323268P.
PR 19-JUN-2002; 2002US-00177488.
PR 26-MAR-2004; 2004US-0557116P.
PR 04-AUG-2004; 2004US-0598899P.
XX
PA (GETH ) GENENTECH INC.
XX
XX Cairns B, Chen R, Frantz G, Hillan KJ, Koeppe H, Phillips HS;
PI Polakis P, Spencer SD, Smith V, Williams PM, Wu TD, Zhang Z;
PI Sakanaka C, Chuntharapai A, Reed CJ;
XX
DR WPI; 2005-371577/38.
DR P-PSDB; AEA00686.
XX
PT Novel isolated antibody e.g. anti-E16 or anti-TAT112 antibody that binds
PT to tumor-associated antigenic target polypeptide, useful for diagnosing
PT or treating cancer.
XX
PS Claim 1; SEQ ID NO 40; 96pp; English.
XX
XX This invention relates to a novel isolated antibody, for example anti-E16
XX or anti-TAT112 antibody, that binds to a tumor-associated antigenic
XX target polypeptide (TAT) and that lacks an associated signal peptide
XX sequence. The invention may be useful for the development of compounds
XX with a cytostatic activity acting as antagonists of the TAT188
XX polypeptide or RNA interference whilst the disclosed sequences may be
XX useful for gene therapy. The invention is useful for inducing the death
XX of a cell (such as a cancer cell chosen from breast, colon, rectum,
XX endometrium, kidney, lung, ovary, skin and liver) to which it binds,
XX inhibiting proliferation or promoting cell death of a cell expressing
XX TAT188. In addition, the invention may be useful for detecting the level
XX of TAT188 polypeptide in a test cell relative to a control cell, or
XX treating or preventing a cell proliferative disorder associated with
XX increased expression of TAT188. The novel antibody of the invention is
XX useful for inhibiting the growth of a cancer cell and may be useful for
XX diagnosing or treating cancer. The present sequence is that of the human
XX TAT115 cDNA which encodes a protein against which an antibody of the
XX invention may be targeted.
XX
SQ Sequence 2456 BP; 519 A; 722 C; 571 G; 644 T; 0 U; 0 Other;
```

```
Query Match 100.0%; Score 17; DB 14; Length 2456;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACGCTTGGCCGAG 17
Db 1271 GCCACGCTTGGCCGAG 1287

RESULT 38
AED47480
ID AED47480 standard; DNA; 2456 BP.
XX
AC AED47480;
XX
DT 15-DEC-2005 (first entry)
XX
DE Retinoic acid-inducible G-protein coupled receptor 3 gene, SEQ ID 2.
XX
KW RNA Interference; gene silencing;
KW retinoic acid-inducible G-protein coupled receptor 3; breast tumor;
KW cytostatic; cell growth; gene; ds.
XX
OS Homo sapiens.
XX
FH Key
FT CDS Location/Qualifiers
FT 254..1327
FT /tag= a
FT /product= "Retinoic acid-inducible G-protein coupled
FT receptor 3"
FT 338..403
FT /tag= b
FT /label= Transmembrane domain
FT 452..514
FT /tag= c
FT /label= Transmembrane domain
FT 497..1057
FT /tag= d
FT /label= 7 transmembrane receptor
FT 548..601
FT /tag= e
FT /label= Transmembrane domain
FT 656..718
FT /tag= f
FT /label= Transmembrane domain
FT 776..832
FT /tag= g
FT /label= Transmembrane domain
FT 887..946
FT /tag= h
FT /label= Transmembrane domain
FT 2403..2408
FT /tag= i
FT polyA_site 2442..2456
FT /tag= j
XX JP2005287456-A.
XX
XX 20-OCT-2005.
XX
XX 02-APR-2004; 2004JP-00110232.
XX
XX 02-APR-2004; 2004JP-00110232.
XX (NIPK ) NIPPON KAYAKU KK.
XX (UYN1-) UNIV NIPPON IKA.
XX
XX Nagahata T, Emi M;
XX WPI; 2005-737982/76.
XX P-PSDB; AED47479.
XX
```

PT Substance that suppresses expression of retinoic acid-inducible G-protein  
PT coupled receptor-3 RAI3 gene or nucleic acid, useful for suppressing cell  
PT growth, and as pharmaceutical or therapeutic agent of breast cancer.  
XX  
XX  
PS Claim 1; SEQ ID NO 2; 22pp; Japanese.  
XX  
CC The invention relates to a novel substance which suppresses the  
CC expression of a retinoic acid-inducible G-protein coupled receptor (RAI)3  
CC gene. The RAI3 gene suppressing agent is useful for suppressing cell  
CC growth, and as a pharmaceutical or therapeutic agent of breast cancer,  
CC and is also useful in identifying a cell growth inhibitory substance. The  
CC RAI3 gene suppressing agent enables the identification of a cell growth  
CC inhibitory substance. This polynucleotide sequence represents the  
CC retinoic acid-inducible G-protein coupled receptor 3 gene of the  
CC invention.  
XX  
SQ Sequence 2456 BP; 519 A; 722 C; 571 G; 644 T; 0 U; 0 Other;  
  
Query Match 100.0%; Score 17; DB 14; Length 2456;  
Best Local Similarity 100.0%; Pred. No. 66;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 GCCCAGCTTGGCCGAG 17  
Db 1271 GCCCAGCTTGGCCGAG 1287  
  
RESULT 39  
ABQ54954  
ID ABQ54954 standard; cDNA; 2593 BP.  
XX  
AC ABQ54954;  
XX  
DT 22-AUG-2002 (first entry)  
XX  
DE Human ovarian antigen HSKJC61 cDNA, SEQ ID NO:834.  
XX  
KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;  
KW ovarian cancer; breast cancer; tumour; reproductive system disorder;  
KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;  
KW PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;  
KW inflammatory condition; immune disorder; blood disorder;  
KW cardiovascular disorder; respiratory disorder; neurological disorder;  
KW gastrointestinal disorder; urinary system disorder; drug screening;  
KW gene therapy; chromosome mapping; forensic analysis;  
KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;  
KW antiinflammatory; gynaecological; reproductive; gene; ss.  
XX  
OS Homo sapiens.  
XX  
WO200200677-A1.  
XX  
PN 03-JAN-2002.  
XX  
PD 07-JUN-2001; 2001WO-US018569.  
EF  
XX 07-JUN-2000; 2000US-0209467P.  
XX  
PR (HUMA-) HUMAN GENOME SCI INC.  
XX  
PA Birse CE, Rosen CA;  
XX  
FI WPI; 2002-147878/19.  
XX  
DR WPI; 2001-375006/39.  
XX  
PT Isolated nucleic acid molecules encoding novel ovarian polypeptides,  
PT useful in the prevention, treatment and diagnosis of cancer (e.g. ovarian  
PT cancer), immune disorders, cardiovascular disorders and neurological  
PT diseases.  
XX  
PS Claim 1; SEQ ID NO 834; 2922pp; English.  
XX  
XX The invention relates to 2175 novel human ovarian antigens (ABP41054-

CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also  
CC encompasses polypeptides 90% identical and polynucleotides 95% identical  
CC to the sequences of the invention. The invention additionally relates to  
CC recombinant vectors and host cells comprising human ovarian antigen  
CC polynucleotides, antibodies against human ovarian antigens, and the use  
CC of ovarian antigen polynucleotides and polypeptides in diagnosing,  
CC treating, prognosing or preventing various ovary and/or breast-related  
CC disorders. Such conditions include ovarian cancer and breast cancer, and  
CC metastatic tumours of ovarian or breast origin, reproductive system  
CC disorders (e.g., infertility, disorders of pregnancy, anovulation,  
CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine  
CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic  
CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and  
CC vaginitis), immune disorders (e.g., congenital and acquired  
CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),  
CC blood-related disorders (e.g., anaemia), cardiovascular disorders,  
CC respiratory disorders, neurological disorders, gastrointestinal disorders  
CC and urinary system disorders. Ovarian antigen polypeptides and  
CC polynucleotides may also be used in screening for compounds which  
CC modulate ovarian antigen expression or activity. The polynucleotides may  
CC further be used for gene therapy, chromosome mapping, in the  
CC identification of individuals and in forensic analysis, and the  
CC polypeptides may be used as food additives or to prepare antibodies  
CC useful in disease diagnosis, drug targeting and phenotyping. The present  
CC sequence represents cDNA encoding a human ovarian antigen of the  
CC invention. Note: The sequence data for this patent did not form part of  
CC the printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 2593 BP; 591 A; 730 C; 605 G; 667 T; 0 U; 0 Other;  
  
Query Match 100.0%; Score 17; DB 6; Length 2593;  
Best Local Similarity 100.0%; Pred. No. 66;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 GCCCAGCTTGGCCGAG 17  
Db 1383 GCCCAGCTTGGCCGAG 1399  
  
RESULT 40  
AAH72766  
ID AAH72766 standard; cDNA; 3371 BP.  
XX  
AC AAH72766;  
XX  
DT 19-SEP-2001 (first entry)  
XX  
DE Human cervical cancer marker nucleic acid 4040.  
XX  
KW Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.  
XX  
OS Homo sapiens.  
XX  
WO200142467-A2.  
XX  
PN 14-JUN-2001.  
FD  
XX 08-DEC-2000; 2000WO-US033312.  
XX  
PR 08-DEC-1999; 99US-0169681P.  
PR 21-DEC-1999; 99US-0171350P.  
PR 14-MAR-2000; 2000US-0189315P.  
PR 12-MAY-2000; 2000US-0203791P.  
PR 09-JUN-2000; 2000US-0210600P.  
PR 21-JUL-2000; 2000US-0220114P.  
XX  
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX  
XX Schlegel R, Deeds J, Berger A, Zhao X;  
XX WPI; 2001-375006/39.  
XX



PT New isolated nucleic acid for diagnosing and treating cervical cancer and  
PT for assessing and detecting compounds for treating the cancer.  
XX  
PS Claim 1; Page 840; 1051pp; English.

CC The invention relates to novel genes (AAH68727-AAH73383) associated with  
CC cervical cancer with cytostatic activity. The nucleic acids and encoded  
CC polypeptides are useful: to assess if a patient is afflicted with  
CC cervical cancer or has a pre-malignant condition; to monitor the  
CC progression of cervical cancer or a premalignant condition in a patient;  
CC and to select and/or assess the efficacy of a compound or therapy for  
CC inhibiting cervical cancer in a patient. The nucleic acids may also be  
CC useful for gene therapy

XX Sequence 3371 BP; 741 A; 902 C; 779 G; 946 T; 0 U; 3 Other;

Query Match 100.0%; Score 17; DB 4; Length 3371;  
Best Local Similarity 100.0%; Pred. No. 66;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCCAGCTTGCCGAG 17  
DB 1184 GCCCAGCTTGCCGAG 1200

RESULT 41  
ACN89274  
ID ACN89274 standard; DNA; 4239 BP.

XX AC ACN89274;

XX 02-DEC-2004 (first entry)

XX Breast cancer related marker, seq id 10424.

XX Cancer; breast; tumour; cytostatic; marker; detection; therapy; ds.

XX Homo sapiens.

XX US2003099974-A1.

XX 29-MAY-2003.

XX 18-JUL-2002; 2002US-00198846.

XX 18-JUL-2001; 2001US-0306220P.

XX (MILL-) MILLENNIUM PHARM INC.

XX Lillie J, Xu Y, Wang Y, Steinmann K;

XX WPI; 2003-787014/74.

XX Novel isolated polypeptide associated with breast cancer, useful for  
XX detecting presence of polypeptide in sample, as a marker for breast  
XX cancer.

XX Disclosure; SEQ ID NO 10424; 36pp; English.

CC The invention relates to an isolated polypeptide (I) associated with  
CC breast cancer which is encoded by a nucleic acid molecule comprising a  
CC nucleotide sequence (S1). Further disclosed is an antibody that binds to  
CC the polypeptide of the invention. The activity of the polypeptide of the  
CC invention may be described as cytostatic. The antibody is useful for  
CC detecting the presence of (I) in a sample. Nucleic acid molecules of the  
CC invention are useful in the detection of breast tumours. (I) is useful as  
CC a marker for breast cancer and in breast cancer therapy. Sequences given  
CC in records ACN78851-ACN92934 represent nucleic acid markers associated  
CC with breast cancer. Note: The sequence listing does not form part of the  
CC specification but may be obtained in electronic format from the USPTO web  
CC site at [seqdata.uspto.gov/sequence.html?DocID=20030099974](http://seqdata.uspto.gov/sequence.html?DocID=20030099974)

XX Sequence 4239 BP; 859 A; 1161 C; 1075 G; 1141 T; 0 U; 3 Other;

Query Match 100.0%; Score 17; DB 11; Length 4239;  
Best Local Similarity 100.0%; Pred. No. 66;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCCAGCTTGCCGAG 17  
DB 1679 GCCCAGCTTGCCGAG 1695

RESULT 42  
ADX97494

ID ADX97494 standard; DNA; 6730 BP.

XX AC ADX97494;

XX 21-APR-2005 (first entry)

XX Pancreatic cancer associated human protein encoding DNA, SEQ ID 42.

XX pancreas tumor; cytostatic; gene; ds.

XX Homo sapiens.

XX EP1471075-A2.

XX 27-OCT-2004.

XX 31-MAR-2004; 2004EP-00090124.

XX 31-MAR-2003; 2003DE-01015834.

XX (HINZ/) HINZMANN B.

XX (ROSE/) ROSENTHAL A.

XX (PILA/) PILARSKY C.

XX (DAHL/) DAHL E.

XX (SPEC/) SPECHT T.

XX (LICH/) LICHTNER R.

XX Rosenthal A, Pilarsky C, Dahl E, Specht T, Bruemendorf T;

XX Lichtner R, Staub E, Roepcke S, Li X;

XX WPI; 2004-768082/76.

XX P-PSDB; ADX97565.

XX New nucleic acid differentially expressed in pancreatic tumor tissue, for  
XX use as diagnostic agents and in screening for therapeutic agents.

XX Claim 1; SEQ ID NO 42; 28pp; German.

CC The invention relates to a novel human nucleic acid sequence of the  
CC pancreas and its encoded protein. The invention further comprises:  
CC proteins and peptides, preferably isolated, that contain a sequence  
CC encoded by the novel nucleic acid; and methods for diagnosis and  
CC treatment of pancreatic cancer, using a substance that inhibits or binds  
CC to the protein or its DNA, including: an antisense oligonucleotide, short  
CC interfering RNA or ribozyme directed against the pancreatic protein, an  
CC organic molecule, particularly having a molecular weight below 5000,  
CC especially 300, that binds to the pancreatic DNA, an aptamer or  
CC (monoclonal) antibody, preferably human or humanized, that binds to the  
CC pancreatic DNA, or an anti-idiotypic antibody raised against the  
CC monoclonal antibody, any of which may be derivatized with a reporter  
CC group, cytotoxic compound, immunostimulant and/or radioisotope. The novel  
CC human pancreatic proteins and their encoding DNA have cytostatic  
CC activity. The novel sequences are useful for inhibiting transcription  
CC and/or expression of genes and proteins associated with pancreatic  
CC cancer. This polynucleotide sequence represents the DNA encoding one of  
CC the novel human pancreatic proteins of the invention. Note: This sequence  
CC is not shown in the specification, it has been electronically downloaded  
CC from a DVD-ROM provided with this specification by the European Patent  
CC Office.

XX Sequence 6730 BP; 1469 A; 1763 C; 1555 G; 1943 T; 0 U; 0 Other;

```
Query Match      100.0%; Score 17; DB 13; Length 6730;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GCCCAGCGTTGGCCGAG 17
      |||||
Db      1271 GCCCAGCGTTGGCCGAG 1287

RESULT 43
ABD12717
ID ABD12717 standard; DNA; 1527 BP.
XX AC ABD12717;
XX XX
DT 29-JUL-2004 (first entry)
DE Pseudomonas aeruginosa polynucleotide #11321.
XX KW Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;
XX KW antibacterial.
XX OS Pseudomonas aeruginosa.
XX PN US6551795-B1.
XX XX
PD 22-APR-2003.
XX XX
PF 18-FEB-1999; 99US-00252991.
XX XX
PR 18-FEB-1998; 98US-0074788P.
XX PR 27-JUL-1998; 98US-0094190P.
XX PA (GENO-) GENOME THERAPEUTICS CORP.
XX PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;
XX WPI; 2003-615309/58.
DR P-PSDB; ABO79146.
XX XX
PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
PT useful as molecular targets for diagnostics, prophylaxis and treatment of
PT pathological conditions resulting from bacterial infection.
XX XX
PS Disclosure; SEQ ID NO 11321; 455pp; English.
XX CC
CC The invention relates to Pseudomonas aeruginosa polypeptides and the
CC polynucleotides encoding them. The sequences are useful in diagnosis and
CC therapy of pathological conditions, as molecular targets for diagnostics,
CC prophylaxis and treatment of pathological conditions resulting from a
CC bacterial infection, for evaluating a compound, such as a polypeptide,
CC for the ability to bind a P. aeruginosa nucleic acid, as components of
CC effective antibacterial targets, as targets for antibacterial drugs,
CC including anti-P. aeruginosa drugs, as templates for recombinant
CC production of P. aeruginosa-derived peptides or polypeptides, as target
CC components for diagnosis and/or treatment of P. aeruginosa-caused
CC infection, and in detection of P. aeruginosa sequences or other sequences
CC of Pseudomonas species using biochip technology. Sequences ABD01397-
CC ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:
CC The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format from USPTO at
CC seqdata.uspto.gov/sequence.html
XX SQ
SQ Sequence 1527 BP; 266 A; 557 C; 475 G; 229 T; 0 U; 0 Other;

Query Match      90.6%; Score 15.4; DB 11; Length 1527;
Best Local Similarity 94.1%; Pred. No. 4.3e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 GCCCAGCGTTGGCCGAG 17
      |||||
Db      25 GCCCAGCGTTGGCCGAG 41

RESULT 45
AEA48186
ID AEA48186 standard; DNA; 2976 BP.
XX AC AEA48186;
XX XX
DT 11-AUG-2005 (first entry)
DE DNA of signal peptide and precursor of serine protease, SEQ ID NO: 2.
XX XX
```

```
RESULT 44
ABQ70591/C
ID ABQ70591 standard; DNA; 2306 BP.
XX AC ABQ70591;
XX XX
DT 29-AUG-2003 (revised)
DT 29-AUG-2002 (first entry)
XX XX
DE Listeria monocytogenes 4b contig DNA sequence #533.
XX KW Antibacterial; Listeria; food contamination; mutational analysis;
XX KW infection; ds.
XX OS Listeria monocytogenes ATCC 19115.
XX PN W0200228891-A2.
XX PD 11-APR-2002.
XX PF 04-OCT-2001; 2001WO-FR003061.
XX PR 04-OCT-2000; 2000FR-00012697.
XX XX
XX (INSP ) INST PASTEUR.
XX PA (CNRS ) CNRS CENT NAT RECH SCI.
XX PI Kunst F, Glaser P;
XX WPI; 2002-332479/37.
DR XX
XX New genomic sequences from Listeria species, useful for detection,
XX treatment and prevention of infection, also related polypeptides,
XX antibodies and modulators.
XX PS Claim 14; SEQ ID NO 3404; 180pp; French.
XX CC
CC The present invention relates to nucleic acid sequences (ABQ67188-
CC ABQ71212) from Listeria sp. The sequences are useful as probes and
CC primers for identification and/or detection of Listeria (e.g. as
CC contaminants in foods, or mutational analysis) and for analysis of gene
CC expression. Proteins encoded by the nucleic acid sequences can be used to
CC screen for compounds that modulate gene expression, replication and
CC pathogenicity of Listeria (potential therapeutic agents), also for
CC treating infections by Listeria, and are useful as immunogens in anti-
CC Listeria vaccines. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences. (Updated
CC on 29-AUG-2003 to standardise OS field)
XX SQ
SQ Sequence 2306 BP; 805 A; 411 C; 489 G; 601 T; 0 U; 0 Other;

Query Match      90.6%; Score 15.4; DB 6; Length 2306;
Best Local Similarity 94.1%; Pred. No. 4.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 GCCCAGCGTTGGCCGAG 17
      |||||
Db      834 GCCCAGCGTTGGCCGAG 818

RESULT 45
AEA48186
ID AEA48186 standard; DNA; 2976 BP.
XX AC AEA48186;
XX XX
DT 11-AUG-2005 (first entry)
DE DNA of signal peptide and precursor of serine protease, SEQ ID NO: 2.
XX XX
```

KW serine protease; hydrolysis; animal feed; detergent; textile; leather;  
 KW ds; gene.  
 XX  
 OS Cellulomonas.  
 XX  
 PN WO2005052161-A2.  
 XX  
 PD 09-JUN-2005.  
 XX  
 PF 19-NOV-2004; 2004WO-US039006.  
 XX  
 PR 19-NOV-2003; 2003US-0523609P.  
 XX  
 PA (GENV ) GENENCOR INT INC.  
 XX  
 PI Jones BE, Kolkman M, Leeflang C, Poulouse AJ, Shaw A;  
 PI Van Der Kleij WA, Van Marrewijk L;  
 XX  
 DR WPI; 2005-425198/43.  
 XX  
 PT New serine protease (isolated from a member of the Micrococccineae) useful  
 PT in e.g. cleaning composition and animal feed composition.  
 XX  
 PS Claim 53; SEQ ID NO 2; 333pp; English.  
 XX  
 CC The present invention relates to new isolated variant serine protease as  
 CC given in SEQ ID NO:8 (I), obtained from a member of the Micrococccineae.  
 CC Also claimed are compositions comprising an isolated serine protease  
 CC having immunological cross-reactivity with the serine proteases obtained  
 CC from a member of the Micrococccineae and particularly the protease  
 CC obtained from Cellulomonas 6984. The variant has improved stability as  
 CC compared to wild-type Cellulomonas 6984 protease. The protease is  
 CC obtained from Cellulomonas, Oerskovia, Cellulosimicrobium,  
 CC Xylanthacterium or Fromicromospora (preferably Cellulomonas 6984). The  
 CC variant serine protease comprises at least one substitution corresponding  
 CC to the amino acid positions in SEQ ID NO. 8, and where the variant  
 CC protease has better performance in at least one property of keratin  
 CC hydrolysis, thermostability, casein activity, LAS stability or cleaning,  
 CC as compared to wild-type Cellulomonas 6984 protease. The invention deals  
 CC with serine proteases, genetic material encoding the proteases,  
 CC proteolytic proteins obtained from Micrococccineae spp, variant proteins  
 CC developed from them, vectors comprising the DNA encoding the protease,  
 CC host cells transformed with the vector DNA and enzymes produced by the  
 CC host cells. (I) is useful in cleaning compositions and animal feed  
 CC compositions, and is useful in laundry and dish detergents. It is useful  
 CC in textile and leather processing compositions. The isolated  
 CC polynucleotide of (I) provides the capability of isolating further  
 CC polynucleotides, which encode proteins having serine protease activity.  
 CC The enzyme compositions have comparable or improved wash performance, as  
 CC compared to presently used subtilisin proteases. The present sequence is  
 CC DNA encoding the signal peptide that is operatively linked to the  
 CC precursor protease of serine protease, SEQ ID NO:2.  
 XX  
 SQ Sequence 2976 BP; 410 A; 1078 C; 1078 G; 410 T; 0 U; 0 Other;  
 Query Match 90.6%; Score 15.4; DB 14; Length 2976;  
 Best Local Similarity 94.1%; Pred. No. 4.4e+02;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GCCACGCTGGCCGAG 17  
 DB 239 GCCACGCTGGCCGAG 255  
 RESULT 46  
 AAX81946  
 ID AAX81946 standard; DNA; 4183 BP.  
 XX  
 AC AAX81946;  
 XX  
 DT 09-SEP-1999 (first entry)  
 XX  
 DE DNA encoding Human nucleotide pyrophosphohydrolase-2.

XX Human; nucleotide pyrophosphohydrolase-2; NTPPH-2; arthropathy;  
 KW Behcet's syndrome; degenerative joint disease; lupus erythematosus;  
 KW rheumatoid arthritis; psoriasis; immunological disease;  
 KW acquired immune deficiency syndrome; Addison's disease;  
 KW ulcerative colitis; osteoporosis; microbial infection; cancer; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9932610-A1.  
 XX  
 PD 01-JUL-1999.  
 XX  
 PF 02-DEC-1998; 98WO-US025558.  
 XX  
 PR 22-DEC-1997; 97US-00996083.  
 XX  
 PA (INCY-) INCYTE PHARM INC.  
 XX  
 PI Magna H, Schaffer P, Lawton M, Yocum SA, Mitchell PG;  
 PI Hutchinson N, Murry LE;  
 XX  
 DR WPI; 1999-418919/35.  
 DR P-PSDB; AAY23754.  
 XX  
 PT Human nucleotide pyrophosphohydrolase-2.  
 XX  
 PS Claim 8; Fig 1A-K; 89pp; English.  
 XX  
 CC The present sequence encodes human nucleotide pyrophosphohydrolase-2  
 CC (NTPPH-2). Antagonists of the protein are used to treat arthropathy (e.g.  
 CC Behcet's syndrome, degenerative joint disease, lupus erythematosus,  
 CC rheumatoid arthritis, psoriasis and many others); immunological disease  
 CC (e.g. acquired immune deficiency syndrome, Addison's disease, ulcerative  
 CC colitis, osteoporosis, microbial infections and many others), also a wide  
 CC range of cancers  
 XX  
 SQ Sequence 4183 BP; 674 A; 1427 C; 1351 G; 728 T; 0 U; 3 Other;  
 Query Match 90.6%; Score 15.4; DB 2; Length 4183;  
 Best Local Similarity 94.1%; Pred. No. 4.4e+02;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GCCACGCTGGCCGAG 17  
 DB 3254 GCCACGCTGGCCGAG 3270  
 RESULT 47  
 ADQ24610  
 ID ADQ24610 standard; DNA; 4222 BP.  
 XX  
 AC ADQ24610;  
 XX  
 DT 26-AUG-2004 (first entry)  
 XX  
 DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 7430.  
 KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;  
 KW ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2004048938-A2.  
 XX  
 PD 10-JUN-2004.  
 XX  
 PF 26-NOV-2003; 2003WO-US038193.  
 XX  
 PR 26-NOV-2002; 2002US-0429739P.  
 XX  
 PA (PROT-) PROTEIN DESIGN LABS INC.  
 XX

```
PI Aziz N, Ginsburg WM, Zlotnik A;
XX WPI; 2004-441208/41.
XX
XX Early detection of soft tissue sarcoma comprises determining expression
PT of a gene in a first soft tissue sample and a normal soft tissue sample
PT and comparing the gene expression, also useful in treating soft tissue
PT sarcoma.
XX
XX Example 2; SEQ ID NO 7430; 210pp; English.
XX
XX The invention relates to a novel method for detecting soft tissue sarcoma
CC which comprises obtaining a first soft tissue sample from an individual
CC and a normal soft tissue sample from the same or different individual,
CC determining the expression of a gene in both samples and comparing the
CC expression of the gene in both soft tissue samples, where a higher level
CC of protein expression in the first soft tissue sample indicates the
CC presence of soft tissue sarcoma. The method of the invention has
CC cytosstatic applications and may be useful for detecting soft tissue
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC acid sequences may be useful in diagnostic and screening applications.
CC The current sequence is that of a human soft tissue sarcoma-upregulated
CC DNA of the invention. The current sequence is not shown within the
CC specification per se but was submitted in CD format by the inventor.
XX
XX Sequence 4222 BP; 672 A; 1411 C; 1348 G; 725 T; 0 U; 66 Other;
SQ
Query Match 90.6%; Score 15.4; DB 12; Length 4222;
Best Local Similarity 94.1%; Pred. No. 4.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 GCCCAGCGTTGGCCGAG 17
Db 3295 GCCCAGCGTTGGCCCAAG 3311
RESULT 48
ADRO7536
ID ADRO7536 standard; cDNA; 4232 BP.
XX
XX ADRO7536;
AC
XX
XX 04-NOV-2004 (first entry)
DT
DE
DE Full length human cDNA useful for treating neurological disease Seq 1042.
XX
XX gene; ss; human; oligo-capping method; diagnostic marker; gene therapy;
XX osteoporosis; neurological disease; Alzheimer's disease;
XX Parkinson's disease; dementia; short memory; cancer;
XX sense or motor function; emotional reaction; fear response; panic;
XX osteopathic; neuroprotective; nootropic; antiparkinsonian; cytostatic;
XX tranquiliser.
XX
XX Homo sapiens.
XX
XX EP1447413-A2.
XX
XX 18-AUG-2004.
XX
XX 12-FEB-2004; 2004EP-00003145.
XX
XX 14-FEB-2003; 2003JP-00102207.
XX
XX 09-MAY-2003; 2003JP-00131452.
XX
XX (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
XX Isogai T, Yamamoto J, Nishikawa T, Isono Y, Sugiyama T, Otsuki T;
PI Wakamatsu A, Ishii S, Nagai K, Irie R;
XX
XX WPI; 2004-583265/57.
XX
XX P-PSDB; ADR09492.
XX
XX New 1995 cDNA, useful for treating osteoporosis, neurological diseases,
PI
```

---

```
PT Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.
XX Claim 1; SEQ ID NO 1042; 2686pp; English.
XX
XX This invention relates to novel, isolated full length human cDNA
CC molecules and the encoded proteins thereof. Specifically, it refers to
CC cDNA clones obtained by an oligo-capping method, where none of these
CC clones are identical to any known human mRNAs. The present invention
CC describes an immunoassay to identify agonists and antagonists, as well as
CC antibodies, antisense molecules and siRNAs that can all be used to bind
CC to and modulate expression of the cDNA molecules. As such, these
CC molecules are useful for diagnostic markers or therapeutic targets for
CC the various diseases or morbid states. In particular, they are useful in
CC gene therapy for treating osteoporosis, neurological disease, Alzheimer's
CC disease, Parkinson's disease, dementia, short memory and various cancers,
CC as well as for maintaining equilibrium of sense or motor function, and
CC for treating emotional reaction, fear response and panic. Accordingly,
CC they exhibit osteopathic, neuroprotective, nootropic, antiparkinsonian,
CC cytosstatic and tranquiliser activities. This polynucleotide is a full
CC length human cDNA sequence of the invention. NOTE: This sequence is not
CC given in the sequence listing of the specification but can be obtained on
CC CD-ROM from the European Patent Office, Vienna Sub-office.
XX
XX Sequence 4232 BP; 665 A; 1460 C; 1371 G; 736 T; 0 U; 0 Other;
SQ
Query Match 90.6%; Score 15.4; DB 13; Length 4232;
Best Local Similarity 94.1%; Pred. No. 4.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 GCCCAGCGTTGGCCGAG 17
Db 3323 GCCCAGCGTTGGCCCAAG 3339
RESULT 49
ABA21144/C
ID ABA21144 standard; DNA; 9699 BP.
XX
XX ABA21144;
AC
XX
XX 23-JAN-2002 (first entry)
DT
DE
DE Human nervous system related polynucleotide SEQ ID NO 13475.
XX
XX Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
XX immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
XX antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
XX antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
XX antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
XX antiparasitic; cardiant; immune disorder; cardiovascular disorder;
XX neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
XX Homo sapiens.
XX
XX WO200159063-A2.
XX
XX 16-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001334.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX
XX 04-FEB-2000; 2000US-0180628P.
XX
XX 24-FEB-2000; 2000US-0184664P.
XX
XX 02-MAR-2000; 2000US-0186350P.
XX
XX 16-MAR-2000; 2000US-0189874P.
XX
XX 17-MAR-2000; 2000US-0190076P.
XX
XX 18-APR-2000; 2000US-0198123P.
XX
XX 19-MAY-2000; 2000US-0205515P.
XX
XX 07-JUN-2000; 2000US-0209467P.
XX
XX 28-JUN-2000; 2000US-0214886P.
XX
XX 30-JUN-2000; 2000US-0215135P.
XX
XX 07-JUL-2000; 2000US-0216647P.
XX
XX 07-JUL-2000; 2000US-0216880P.
```

PR 11-JUL-2000; 2000US-02117487P.  
PR 11-JUL-2000; 2000US-02117496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226688P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235835P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 20-OCT-2000; 2000US-0241827P.  
PR 20-OCT-2000; 2000US-0242221P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 01-DEC-2000; 2000US-0251160P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-541565/60.  
XX  
PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating nervous system cancers  
PT and metastases.  
XX  
PS Disclosure; SEQ ID NO 13475; 1701pp + Sequence Listing; English.  
XX  
CC The invention relates to novel genes (ABA1004-ABA21534) and proteins  
CC (ABBI4678-ABBI8001) useful for preventing, treating or ameliorating  
CC medical conditions e.g. by protein or gene therapy. The genes are  
CC isolated from a range of human tissues disclosed in the specification.  
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in  
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and  
CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,  
CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune  
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic  
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,

```
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC cardiovascular disorders such as myocardial ischaemias; (d) wound healing
CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC infectious diseases such as viral, bacterial, fungal and parasitic
CC infections. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 9699 BP; 1831 A; 2869 C; 3150 G; 1849 T; 0 U; 0 Other;
Query Match 90.6%; Score 15.4; DB 5; Length 9699;
Best Local Similarity 94.1%; Pred. No. 4.5e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 GCCCAGCGTTGGCGGAG 17
Db 678 GCCCAGCGTTGGCGGAG 662
RESULT 50
ABA21143/c
ID ABA21143 standard; DNA; 17457 BP.
XX
AC ABA21143;
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polynucleotide SEQ ID NO 13474.
XX
KW Human; nontropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antinflammatory; anti-HIV; antibacterial; vulnery;
KW antiparkinsonian; antieckling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS Homo sapiens.
XX
PN WO200159063-A2.
XX
PD 16-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001334.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232377P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 21-SEP-2000; 2000US-0233085P.
PR 21-SEP-2000; 2000US-0234223P.
PR 25-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 20-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 20-OCT-2000; 2000US-0242221P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
```

PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246612P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249219P.  
PR 17-NOV-2000; 2000US-0249220P.  
PR 17-NOV-2000; 2000US-0249221P.  
PR 17-NOV-2000; 2000US-0249222P.  
PR 17-NOV-2000; 2000US-0249223P.  
PR 17-NOV-2000; 2000US-0249224P.  
PR 17-NOV-2000; 2000US-0249225P.  
PR 17-NOV-2000; 2000US-0249226P.  
PR 17-NOV-2000; 2000US-0249227P.  
PR 17-NOV-2000; 2000US-0249228P.  
PR 17-NOV-2000; 2000US-0249229P.  
PR 17-NOV-2000; 2000US-0249230P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 01-DEC-2000; 2000US-0251160P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX PA  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX PI Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-541565/60.  
XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating nervous system cancers  
PT and metastases.  
XX FS Disclosure; SEQ ID NO 13474; 1701pp + Sequence Listing; English.  
XX CC The invention relates to novel genes (AB11004-AB21534) and proteins  
CC (AB114678-AB18001) useful for preventing, treating or ameliorating  
CC medical conditions e.g. by protein or gene therapy. The genes are  
CC isolated from a range of human tissues disclosed in the specification.  
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in  
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and  
CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,  
CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune  
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic  
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,  
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)  
CC cardiovascular disorders such as myocardial ischaemia; (d) wound healing  
CC; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)  
CC infectious diseases such as viral, bacterial, fungal and parasitic  
CC infections. Note: The sequence data for this patent did not form part of  
CC the printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX SQ Sequence 17457 BP; 3531 A; 4969 C; 5360 G; 3597 T; 0 U; 0 Other;  
Query Match 90.6%; Score 15.4; DB 5; Length 17457;  
Best Local Similarity 94.1%; Fred. NO. 4.6e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Oy 1 GCCCAGCTTGGCGAG 17

Db 8434 GCCCAGCTTGGCGAG 8418  
RESULT 51  
ABA21142/c  
ID ABA21142 standard; DNA; 17458 BP.  
XX AC ABA21142;  
XX DT 23-JAN-2002 (first entry)  
XX DE Human nervous system related polynucleotide SEQ ID NO 13473.  
XX KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;  
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;  
KW antiparkinsonian; antiskilling; antianaemic; antiarthritic; cancer;  
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;  
KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;  
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.  
XX OS Homo sapiens.  
XX PN WO200159063-A2.  
XX PD 16-AUG-2001.  
XX PF 17-JAN-2001; 2001WO-US001334.  
XX PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 11-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 14-JUL-2000; 2000US-0217496P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225477P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 18-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 23-AUG-2000; 2000US-0227182P.  
PR 30-AUG-2000; 2000US-0227009P.  
PR 01-SEP-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.

PR	06-SEP-2000;	2000US-02304338P
PR	08-SEP-2000;	2000US-02312422P
PR	08-SEP-2000;	2000US-02312433P
PR	08-SEP-2000;	2000US-02312444P
PR	08-SEP-2000;	2000US-0231413P
PR	08-SEP-2000;	2000US-0231414P
PR	08-SEP-2000;	2000US-0232080P
PR	08-SEP-2000;	2000US-0232081P
PR	12-SEP-2000;	2000US-0231968P
PR	14-SEP-2000;	2000US-0233197P
PR	14-SEP-2000;	2000US-0233298P
PR	14-SEP-2000;	2000US-0233299P
PR	14-SEP-2000;	2000US-023400P
PR	14-SEP-2000;	2000US-0232401P
PR	14-SEP-2000;	2000US-0233603P
PR	14-SEP-2000;	2000US-0233064P
PR	14-SEP-2000;	2000US-0233065P
PR	21-SEP-2000;	2000US-0234223P
PR	21-SEP-2000;	2000US-0234274P
PR	25-SEP-2000;	2000US-0234997P
PR	25-SEP-2000;	2000US-0234998P
PR	26-SEP-2000;	2000US-023484P
PR	27-SEP-2000;	2000US-0233834P
PR	27-SEP-2000;	2000US-0233836P
PR	27-SEP-2000;	2000US-0236327P
PR	29-SEP-2000;	2000US-0236367P
PR	29-SEP-2000;	2000US-0236368P
PR	29-SEP-2000;	2000US-0236369P
PR	29-SEP-2000;	2000US-0236370P
PR	29-SEP-2000;	2000US-0236802P
PR	02-OCT-2000;	2000US-0236802P
PR	02-OCT-2000;	2000US-0237037P
PR	02-OCT-2000;	2000US-0237038P
PR	02-OCT-2000;	2000US-0237039P
PR	02-OCT-2000;	2000US-0237040P
PR	13-OCT-2000;	2000US-0239933P
PR	13-OCT-2000;	2000US-0239937P
PR	20-OCT-2000;	2000US-0240960P
PR	20-OCT-2000;	2000US-0241785P
PR	20-OCT-2000;	2000US-0241786P
PR	20-OCT-2000;	2000US-0241787P
PR	20-OCT-2000;	2000US-0241808P
PR	20-OCT-2000;	2000US-0241809P
PR	08-NOV-2000;	2000US-0246475P
PR	08-NOV-2000;	2000US-0246476P
PR	08-NOV-2000;	2000US-0246477P
PR	08-NOV-2000;	2000US-0246478P
PR	08-NOV-2000;	2000US-0246523P
PR	08-NOV-2000;	2000US-024524P
PR	08-NOV-2000;	2000US-024525P
PR	08-NOV-2000;	2000US-024526P
PR	08-NOV-2000;	2000US-024527P
PR	08-NOV-2000;	2000US-0246513P
PR	17-NOV-2000;	2000US-0249207P
PR	17-NOV-2000;	2000US-0249208P
PR	17-NOV-2000;	2000US-0249212P
PR	17-NOV-2000;	2000US-0249213P
PR	17-NOV-2000;	2000US-0249214P
PR	17-NOV-2000;	2000US-0249215P
PR	17-NOV-2000;	2000US-0249216P
PR	17-NOV-2000;	2000US-0249217P
PR	17-NOV-2000;	2000US-0249218P

PR	17-NOV-2000;	2000US-0249244P.
PR	17-NOV-2000;	2000US-0249245P.
PR	17-NOV-2000;	2000US-0249264P.
PR	17-NOV-2000;	2000US-0249265P.
PR	17-NOV-2000;	2000US-0249297P.
PR	17-NOV-2000;	2000US-0249299P.
PR	17-NOV-2000;	2000US-0249300P.
PR	01-DEC-2000;	2000US-0250391P.
PR	01-DEC-2000;	2000US-0251160P.
PR	05-DEC-2000;	2000US-0251030P.
PR	05-DEC-2000;	2000US-0251988P.
PR	05-DEC-2000;	2000US-0256719P.
PR	06-DEC-2000;	2000US-0251479P.
PR	08-DEC-2000;	2000US-0251856P.
PR	08-DEC-2000;	2000US-0251868P.
PR	08-DEC-2000;	2000US-0251869P.
PR	08-DEC-2000;	2000US-0251989P.
PR	08-DEC-2000;	2000US-0251990P.
PR	11-DEC-2000;	2000US-0254097P.
PR	05-JAN-2001;	2001US-0259678P.
XX		
PA	(HUMA-) HUMAN GENOME SCI INC.	
XX		
PI	Rosen CA, Barash SC, Ruben SM;	
XX		
DR	WPI; 2001-541565/60.	
XX		
PT	Nucleic acids encoding 3224 human nervous system antigen polypeptides,	
PT	useful for preventing, diagnosing and/or treating nervous system cancers	
FT	and metastases.	
XX		
PS	Disclosure; SEQ ID NO 13473; 1701pp + Sequence Listing; English.	
XX		
CC	The invention relates to novel genes (ABAI1004-ABA21534) and proteins	
CC	(ABB14678-ABB19001) useful for preventing, treating or ameliorating	
CC	medical conditions e.g. by protein or gene therapy. The genes are	
CC	isolated from a range of human tissues disclosed in the specification..	
CC	The nucleic acids, proteins, antibodies and antagonists are useful in	
CC	the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and	
CC	ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,	
CC	breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune	
CC	disorders e.g. Addison's disease, allergies, autoimmune haemolytic	
CC	anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,	
CC	multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)	
CC	cardiovascular disorders such as myocardial ischaemias; (d) wound healing	
CC	; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)	
CC	infectious diseases such as viral, bacterial, fungal and parasitic	
CC	infections. Note: the sequence data for this patent did not form part of	
CC	the printed specification, but was obtained in electronic format directly	
CC	from WIPO at ftp.wipo.int/pub/published_pct_sequences	
XX		
SQ	Sequence 17458 BP; 3531 A; 4569 C; 5361 G; 3597 T; 0 U; 0 Other;	
	Query Match 90.6%; Score 15.4; DB 5; Length 17458;	
	Best Local Similarity 94.1%; Pred.No. 4.6e+02;	
	Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
Qy	1 GCCCACGCTTGGCGGAG 17	
Db	8434 GCCCACGCTTGGCGGAG 8418	
RESULT 52		
ABD33426		
ID ABD33426 standard; DNA; 63761 BP.		
XX		
AC ABD33426;		
XX		
DT 18-NOV-2004 (first entry)		
XX		
DE Murine cancer-associated (CA) gene MD07-080.		
XX		
KW Mouse; cancer-associated protein; CAP; cancer-associated gene; CA; gene;		



KW ds; cancer; cytostatic.  
XX Mus musculus.  
XX WO2004058146-A2.  
XX 15-JUL-2004.  
XX 15-DEC-2003; 2003WO-US040081.  
XX 17-DEC-2002; 2002US-00322281.  
XX (SAGR-) SAGRES DISCOVERY INC.  
XX Morris DW, Malandro MS;  
XX WPI; 2004-499109/47.  
XX Novel human cancer associated protein encoded within open reading frame  
XX of cancer associated gene, useful as targets for diagnosing cancer.  
XX Disclosure; SEQ ID NO 557; 182pp; English.  
XX The invention relates to cancer-associated proteins (CAP) and the cancer-  
XX associated (CA) nucleic acids encoding them. The invention also relates  
XX to a method for treating cancers involving administering to a patient an  
XX inhibitor of CAP, and a method of screening for anticancer activity in a  
XX potential drug involving providing a cell that expresses a CA gene,  
XX contacting a tissue sample derived from a cancer cell with an anticancer  
XX drug candidate and monitoring the effect of the anticancer drug candidate  
XX on expression of the CA gene. The CAP proteins are useful for detecting  
XX cancer associated with expression of a CAP protein in a test cell sample  
XX and for screening for a bioactive agent capable of modulating the  
XX activity of a CAP protein. The CA nucleic acids are useful for diagnosing  
XX cancer, involving determining the expression of a CA nucleic acid in a  
XX tissue. This sequence represents a murine CA gene of the invention. Note:  
XX The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 63761 BP; 13712 A; 15513 C; 16452 G; 15636 T; 0 U; 2448 Other;  
XX  
XX Query Match 90.6%; Score 15.4; DB 13; Length 63761;  
XX Best Local Similarity 94.1%; Pred. No. 4.7e+02;  
XX Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
XX  
XX QY 1 GCCACGCTTGCCGAG 17  
XX |||||  
XX Db 34522 GCCACGCTTGCCGAG 34538  
XX  
XX RESULT 53  
XX ABD33428  
XX ID ABD33428 standard; DNA; 68732 BP.  
XX AC ABD33428;  
XX 18-NOV-2004 (first entry)  
XX Human cancer-associated (CA) gene HD07-080.  
XX Human; cancer-associated protein; CAP; cancer-associated gene; CA; gene;  
XX ds; cancer; cytostatic.  
XX Homo sapiens.  
XX WO2004058146-A2.  
XX 15-JUL-2004.  
XX 15-DEC-2003; 2003WO-US040081.  
XX 17-DEC-2002; 2002US-00322281.

XX (SAGR-) SAGRES DISCOVERY INC.  
XX Morris DW, Malandro MS;  
XX WPI; 2004-499109/47.  
XX Novel human cancer associated protein encoded within open reading frame  
XX of cancer associated gene, useful as targets for diagnosing cancer.  
XX Claim 16; SEQ ID NO 560; 182pp; English.  
XX The invention relates to cancer-associated proteins (CAP) and the cancer-  
XX associated (CA) nucleic acids encoding them. The invention also relates  
XX to a method for treating cancers involving administering to a patient an  
XX inhibitor of CAP, and a method of screening for anticancer activity in a  
XX potential drug involving providing a cell that expresses a CA gene,  
XX contacting a tissue sample derived from a cancer cell with an anticancer  
XX drug candidate and monitoring the effect of the anticancer drug candidate  
XX on expression of the CA gene. The CAP proteins are useful for detecting  
XX cancer associated with expression of a CAP protein in a test cell sample  
XX and for screening for a bioactive agent capable of modulating the  
XX activity of a CAP protein. The CA nucleic acids are useful for diagnosing  
XX cancer, involving determining the expression of a CA nucleic acid in a  
XX tissue. This sequence represents a human CA gene of the invention. Note:  
XX The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 68732 BP; 14866 A; 17372 C; 18480 G; 16597 T; 0 U; 1417 Other;  
XX  
XX Query Match 90.6%; Score 15.4; DB 13; Length 68732;  
XX Best Local Similarity 94.1%; Pred. No. 4.7e+02;  
XX Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
XX  
XX QY 1 GCCACGCTTGCCGAG 17  
XX |||||  
XX Db 38569 GCCACGCTTGCCGAG 38585  
XX  
XX RESULT 54  
XX ABQ67195\_4  
XX Continuation (5 of 5) of ABQ67195 from base 400001 (Listeria innocua contig DNA sequence  
XX WP Sequence split into 5 fragments LOCUS ABQ67195 Accession Abq67195  
XX WP Fragment Name Begin End  
XX WP ABQ67195\_0 1 110000  
XX WP ABQ67195\_1 100001 210000  
XX WP ABQ67195\_2 200001 310000  
XX WP ABQ67195\_3 300001 410000  
XX WP ABQ67195\_4 400001 495269  
XX  
XX Query Match 90.6%; Score 15.4; DB 6; Length 95269;  
XX Best Local Similarity 94.1%; Pred. No. 4.7e+02;  
XX Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
XX  
XX QY 1 GCCACGCTTGCCGAG 17  
XX |||||  
XX Db 813 GCCACGCTTGCTGAG 829  
XX  
XX RESULT 55  
XX ABQ69245\_13/c  
XX Continuation (14 of 31) of ABQ69245 from base 1300001 (Listeria innocua DNA sequence #68  
XX WP Sequence split into 31 fragments LOCUS ABQ69245 Accession Abq69245  
XX WP Fragment Name Begin End  
XX WP ABQ69245\_00 1 110000  
XX WP ABQ69245\_01 100001 210000  
XX WP ABQ69245\_02 200001 310000  
XX WP ABQ69245\_03 300001 410000  
XX WP ABQ69245\_04 400001 510000  
XX WP ABQ69245\_05 500001 610000  
XX WP ABQ69245\_06 600001 710000  
XX WP ABQ69245\_07 700001 810000

WP ABQ69245\_08 800001 910000  
WP ABQ69245\_09 900001 1010000  
WP ABQ69245\_10 100001 1110000  
WP ABQ69245\_11 1100001 1210000  
WP ABQ69245\_12 1200001 1310000  
WP ABQ69245\_13 1300001 1410000  
WP ABQ69245\_14 1400001 1510000  
WP ABQ69245\_15 1500001 1610000  
WP ABQ69245\_16 1600001 1710000  
WP ABQ69245\_17 1700001 1810000  
WP ABQ69245\_18 1800001 1910000  
WP ABQ69245\_19 1900001 2010000  
WP ABQ69245\_20 2000001 2110000  
WP ABQ69245\_21 2100001 2210000  
WP ABQ69245\_22 2200001 2310000  
WP ABQ69245\_23 2300001 2410000  
WP ABQ69245\_24 2400001 2510000  
WP ABQ69245\_25 2500001 2610000  
WP ABQ69245\_26 2600001 2710000  
WP ABQ69245\_27 2700001 2810000  
WP ABQ69245\_28 2800001 2910000  
WP ABQ69245\_29 2900001 3010000  
WP ABQ69245\_30 3000001 3011208

Query Match 90.6%; Score 15.4; DB 6; Length 110000;  
Best Local Similarity 94.1%; Pred. No. 4.8e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17  
|||||

Db 58141 GCCCAGCGTTGGCTGAG 58125

## RESULT 56

ABQ67195\_3  
Continuation (4 of 5) of ABQ67195 from base 300001 (Listeria innocua contig DNA sequence  
WP Sequence split into 5 fragments LOCUS ABQ67195 Accession Abq67195

WP Fragment Name Begin End  
WP ABQ67195\_0 1 110000  
WP ABQ67195\_1 100001 210000  
WP ABQ67195\_2 200001 310000  
WP ABQ67195\_3 300001 410000  
WP ABQ67195\_4 400001 495269

Query Match 90.6%; Score 15.4; DB 6; Length 110000;  
Best Local Similarity 94.1%; Pred. No. 4.8e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17  
|||||

Db 100813 GCCCAGCGTTGGCTGAG 100829

## RESULT 57

ABA03041\_13/c  
Continuation (14 of 30) of ABA03041 from base 1300001 (Listeria monocytogenes EGD-e gene  
WP Sequence split into 30 fragments LOCUS ABA03041 Accession ABA03041

WP Fragment Name Begin End  
WP ABA03041\_00 1 110000  
WP ABA03041\_01 100001 210000  
WP ABA03041\_02 200001 310000  
WP ABA03041\_03 300001 410000  
WP ABA03041\_04 400001 510000  
WP ABA03041\_05 500001 610000  
WP ABA03041\_06 600001 710000  
WP ABA03041\_07 700001 810000  
WP ABA03041\_08 800001 910000  
WP ABA03041\_09 900001 1010000  
WP ABA03041\_10 1000001 1110000  
WP ABA03041\_11 1100001 1210000  
WP ABA03041\_12 1200001 1310000  
WP ABA03041\_13 1300001 1410000  
WP ABA03041\_14 1400001 1510000

WP ABA03041\_15 1500001 1610000  
WP ABA03041\_16 1600001 1710000  
WP ABA03041\_17 1700001 1810000  
WP ABA03041\_18 1800001 1910000  
WP ABA03041\_19 1900001 2010000  
WP ABA03041\_20 2000001 2110000  
WP ABA03041\_21 2100001 2210000  
WP ABA03041\_22 2200001 2310000  
WP ABA03041\_23 2300001 2410000  
WP ABA03041\_24 2400001 2510000  
WP ABA03041\_25 2500001 2610000  
WP ABA03041\_26 2600001 2710000  
WP ABA03041\_27 2700001 2810000  
WP ABA03041\_28 2800001 2910000  
WP ABA03041\_29 2900001 2944528

Query Match 90.6%; Score 15.4; DB 6; Length 110000;  
Best Local Similarity 94.1%; Pred. No. 4.8e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17  
|||||

Db 51366 GCCCAGCGTTGGCTGAG 51350

## RESULT 58

ABQ81846/c

ID ABQ81846 standard; DNA; 349980 BP.

XX AC ABQ81846;

XX DT 19-NOV-2002 (first entry)

XX DE Bifidobacterium longum NCC2705 related nucleotide sequence SEQ ID:1102.

XX KW Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;

XX KW antidiarrheic; antibacterial; inhibitor of Salmonella; detection;

XX KW identification; lactic acid bacterium; diarrhoea; pathogenic bacteria;  
XX KW rotavirus; food composition; pharmaceutical composition; gene; ds.

XX OS Bifidobacterium longum.

XX OS Synthetic.

XX EN EPI227152-A1.

XX PD 31-JUL-2002.

XX PF 30-JAN-2001; 2001EP-00102050.

XX PR 30-JAN-2001; 2001EP-00102050.

XX PA (NEST ) SOC PROD NESTLE SA.

XX DR WPI; 2002-668397/72.

XX PT Novel polynucleotide comprising Bifidobacterium genome sequence useful as  
XX PT a probe or primer for detecting and/or identifying Bifidobacterium longum  
XX PT in a biological sample.

XX PS Disclosure; SEQ ID NO 1102; 80pp; English.

XX CC The present invention describes a polynucleotide (I) comprising a  
XX CC sequence of a Bifidobacterium genome selected from the nucleotide  
XX CC sequences given in ABQ81842 and ABQ81843, or a sequence exhibiting at  
XX CC least 90% identity or which hybridises with the sequences given in  
XX CC ABQ81842 and ABQ81843. Also described is a polynucleotide (II) encoding a  
XX CC fusion protein, comprising a sequence selected from 1097 sequences given  
XX CC in ABP6528 to ABP6354 ligated in frame to a polynucleotide encoding a  
XX CC heterologous polypeptide. (I) has antidiarrheic and antibacterial  
XX CC activities, and can be used as an inhibitor of Salmonella. (I) (which is  
XX CC a probe) is useful for the detection and/or identification of  
XX CC Bifidobacterium longum in a biological sample. A carrier containing the  
XX CC lactic acid bacterium Bifidobacterium longum NCC2705 (CNCM I-2618) can be

used for preventing and/or treating diarrhoea brought about by pathogenic bacteria and/or rotavirus. The carrier is a food composition selected from milk, yogurt, curd, cheese, fermented milks, milk based fermented products, ice-creams, fermented cereal based products, milk based powders, infant formula, pet food or a pharmaceutical composition selected from tablets, liquid bacterial suspensions, dried oral supplement, wet oral supplement, dry tube feeding or wet tube feeding. (I) is useful in DNA arrays or chips to carry out analysis of the expression of the Bifidobacterium gene. ABQ81844 to ABQ81850 represent Bifidobacterium related nucleotide sequences given in the Sequence Listing from the present invention but not mentioned further within the specification. N.B. The sequence data for this patent is not represented in the printed specification but is based on sequence information supplied by the European Patent Office

Sequence 349980 BP; 69200 A; 103414 C; 105147 G; 72219 T; 0 U; 0 Other;

Query Match 90.6%; Score 15.4; DB 6; Length 349980;

Best Local Similarity 94.1%; Pred. No. 4.8e+02;

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCCAGCTTGCCGAG 17

DB 322894 GCCCAGCTTGCCGAG 322878

#### RESULT 59

ABQ81847/c

ID ABQ81847 standard; DNA; 349980 BP.

AC ABQ81847;

DT 19-NOV-2002 (first entry)

XX Bifidobacterium longum NCC2705 related nucleotide sequence SEQ ID:1103.

DE Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;  
KW antidiarrheic; antibacterial; inhibitor of Salmonella; detection;  
KW identification; lactic acid bacterium; diarrhoea; pathogenic bacteria;  
KW rotavirus; food composition; pharmaceutical composition; gene; ds.

XX Bifidobacterium longum.

OS Synthetic.

XX EPI227152-A1.

XX 31-JUL-2002.

XX 30-JAN-2001; 2001EP-00102050.

XX 30-JAN-2001; 2001EP-00102050.

XX (NEST ) SOC PROD NESTLE SA.

XX WPI; 2002-668397/72.

DR Novel polynucleotide comprising Bifidobacterium genome sequence useful as a probe or primer for detecting and/or identifying Bifidobacterium longum in a biological sample.

PS Disclosure; SEQ ID NO 1103; 80pp; English.

XX The present invention describes a polynucleotide (I) comprising a sequence of a Bifidobacterium genome selected from the nucleotide sequences given in ABQ81842 and ABQ81843, or a sequence exhibiting at least 90% identity or which hybridises with the sequences given in ABQ81842 and ABQ81843. Also described is a polynucleotide (II) encoding a fusion protein, comprising a sequence selected from 1097 sequences given in ABP55258 to ABP6354 ligated in frame to a polynucleotide encoding a heterologous polypeptide. (I) has antidiarrheic and antibacterial activities, and can be used as an inhibitor of Salmonella. (I) (which is a probe) is useful for the detection and/or identification of Bifidobacterium longum in a biological sample. A carrier containing the

CC lactic acid bacterium Bifidobacterium longum NCC2705 (NCIM 1-2618) can be used for preventing and/or treating diarrhoea brought about by pathogenic bacteria and/or rotavirus. The carrier is a food composition selected from milk, yogurt, curd, cheese, fermented milks, milk based fermented products, ice-creams, fermented cereal based products, milk based powders, infant formula, pet food or a pharmaceutical composition selected from tablets, liquid bacterial suspensions, dried oral supplement, wet oral supplement, dry tube feeding or wet tube feeding. (I) is useful in DNA arrays or chips to carry out analysis of the expression of the Bifidobacterium gene. ABQ81844 to ABQ81850 represent Bifidobacterium related nucleotide sequences given in the Sequence Listing from the present invention but not mentioned further within the specification. N.B. The sequence data for this patent is not represented in the printed specification but is based on sequence information supplied by the European Patent Office

Sequence 349980 BP; 69124 A; 105769 C; 105545 G; 69542 T; 0 U; 0 Other;

Query Match 90.6%; Score 15.4; DB 6; Length 349980;

Best Local Similarity 94.1%; Pred. No. 4.8e+02;

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCCAGCTTGCCGAG 17

DB 22894 GCCCAGCTTGCCGAG 22878

#### RESULT 60

ABV57592/c

ID ABV57592 standard; cDNA; 337 BP.

AC ABV57592;

DT 17-SEP-2002 (first entry)

XX Human prostate expression marker cDNA 57583.

DE Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
KW pharmacogenomic marker; gene; ss.

XX Homo sapiens.

XX WO200160860-A2.

XX 23-AUG-2001.

XX 20-FEB-2001; 2001WO-US005171.

XX 17-FEB-2000; 2000US-0183319P.

XX 16-MAR-2000; 2000US-0189862P.

XX 25-MAY-2000; 2000US-0207454P.

XX 09-JUN-2000; 2000US-0211314P.

XX 18-JUL-2000; 2000US-0219007P.

XX 13-DEC-2000; 2000US-0255281P.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX Schlegel R, Endege WO, Monahan JE;

XX WPI; 2001-662795/76.

XX Novel isolated nucleic acid molecule associated with cancerous state of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer.

XX Claim 1; Page 11075; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the specification or its complement. (I) is useful for: (a) assessing whether a patient is afflicted with prostate cancer; (b) monitoring the progression of prostate cancer in a patient; (c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient; (d) assessing

CC the efficacy of a therapy for inhibiting prostate cancer in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)  
CC determining whether prostate cancer has metastasized in a patient; (h)  
CC assessing the aggressiveness or indolence of prostate cancer in a patient  
CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker  
XX  
SQ Sequence 337 BP; 88 A; 101 C; 60 G; 88 T; 0 U; 0 Other;  
  
Query Match 88.2%; Score 15; DB 5; Length 337;  
Best Local Similarity 100.0%; Pred. No. 6.7e+02;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 3 CCACGCTTGGCCGAG 17  
Db 216 CCACGCTTGGCCGAG 202  
|||||  
  
RESULT 61  
ADT19568/c  
ID ADT19568 standard; cDNA; 3303 BP.  
XX  
AC ADT19568;  
DT 13-JAN-2005 (first entry)  
XX  
DE Plant cDNA, Seq ID 4894.  
XX  
KW Plant; ss; gene; transgenic; cold tolerance; growth rate;  
KW drought tolerance; disease resistance; galactomannan production;  
KW plant growth regulator; heat tolerance; herbicide tolerance;  
KW lignin production; extreme osmotic condition tolerance;  
KW pathogens resistance; pest resistance; yield improvement; seed oil yield;  
KW seed protein yield.  
XX  
OS Viridiplantae.  
XX  
PN US2004216190-A1.  
XX  
XX 28-OCT-2004.  
XX  
XX 18-DEC-2003; 2003US-00739930.  
XX  
XX 28-APR-2003; 2003US-00424599.  
XX  
XX 28-APR-2003; 2003US-00425115.  
XX  
XX (KOVA/) KOVALIC D K.  
XX  
XX Kovalic DK;  
XX  
XX WPI; 2004-757369/74.  
XX  
XX New recombinant DNA constructs useful in the field of biochemistry and  
XX genetics, and in particular for producing transgenic plants with improved  
XX biological characteristics.  
XX  
XX Claim 1; SEQ ID NO 4894; 14pp; English.  
XX  
XX The invention relates a recombinant DNA construct comprising a  
XX polynucleotide having any of 5544 nucleotide sequences (cDNAs SEQ ID NO:  
XX 1-5544) and encoding a polypeptide with any of 5544 amino acid sequences  
XX (SEQ ID NO: 5545-11088). The cDNAs and proteins are from corn, soybean,  
XX Arabidopsis, wheat and rape but the specification does not indicate which  
XX sequences is derived from which organism. Also included is a method of  
XX producing a plant having an improved property, comprising transforming a  
XX plant with a recombinant DNA construct comprising a promoter region  
XX functional in a plant cell operably joined to a polynucleotide encoding a  
XX polypeptide associated with the property, and growing the transformed  
XX plant. The property is selected from improving plant cold tolerance, for  
XX manipulating growth rate in plant cells by modification of the cell cycle  
XX pathway, for improving plant drought tolerance, for providing increased  
XX resistance to plant disease, for galactomannan production, for production  
XX of plant growth regulators, for improving plant heat tolerance, for

CC improving plant tolerance to herbicides, for increasing the rate of  
CC homologous recombination in plants, for lignin production, for improving  
CC plant tolerance to extreme osmotic conditions, for improving plant  
CC tolerance to pathogens or pests, for yield improvement by modification of  
CC photosynthesis, for modifying seed oil yield and/or content, for  
CC modifying seed protein yield and/or content, for yield improvement by  
CC modification of carbohydrate, nitrogen or phosphorus use and/or uptake  
CC and for yield improvement by providing improved plant growth and  
CC development under at least one stress condition. The polynucleotide may  
CC also encode a plant transcription factor. The methods and compositions of  
CC the present invention are useful in the field of biochemistry and  
CC genetics, in particular for producing transgenic plants with improved  
CC biological characteristics such as increased yield, improved plant  
CC flow, increasing plant tolerance to cold or heat, improving plant  
CC tolerance to extreme osmotic and drought conditions, and improving plant  
CC tolerance to plant pests or pathogens. They can also be used in physical  
CC arrays of molecules, plant breeding markers, computer-based storage and  
CC analysis systems. The present sequence is one of the 5544 plant cDNA  
CC sequences of the invention. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from USPTO at  
CC segdata.uspto.gov/sequence.html?docID=20040216190.  
XX  
SQ Sequence 3303 BP; 637 A; 1063 C; 991 G; 612 T; 0 U; 0 Other;  
  
Query Match 88.2%; Score 15; DB 13; Length 3303;  
Best Local Similarity 100.0%; Pred. No. 7.1e+02;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 2 CCACACGCTTGGCCGA 16  
Db 619 CCACACGCTTGGCCGA 605  
|||||  
  
RESULT 62  
AAS37371/c  
ID AAS37371 standard; cDNA; 344 BP.  
XX  
AC AAS37371;  
XX  
XX 17-DEC-2001 (first entry)  
XX  
DE Novel human diagnostic and therapeutic gene #429.  
XX  
XX Human; cancer; breast; lung; colon; prostate; cytostatic; diagnostic; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200166753-A2.  
XX  
XX 13-SEP-2001.  
XX  
XX 09-MAR-2001; 2001WO-US007787.  
XX  
XX 09-MAR-2000; 2000US-0188609P.  
XX  
XX (CHIR ) CHIRON CORP.  
XX (HYSE-) HYSEQ INC.  
XX  
XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;  
XX Reinhard C, Randazzo F, Kennedy GC, Pot D, Kassam A, Lamson G;  
XX Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;  
XX Leshkowitz D, Kita D, Garcia V, Jones WL, Stache-Crain B;  
XX WPI; 2001-530177/58.  
XX  
XX New polynucleotides and polypeptides, useful for diagnosis and treatment  
XX of breast, lung and colon cancer.  
XX  
XX Claim 1; Page 704; 1193pp; English.  
XX  
XX The invention relates to new polynucleotides and polypeptides, useful for  
XX diagnosis and treatment of breast, lung and colon cancer. The sequences

CC can be used in detecting differentially expressed genes correlated with a  
CC cancerous state of a mammalian cell, comprising detecting at least one  
CC differentially expressed gene product in a test sample derived from a  
CC cell suspected of being cancerous. They can also be used to inhibit  
CC tumour growth by modulating expression of a gene product. AAS36943-  
CC AAS39338 represent novel human diagnostic and therapeutic coding  
CC sequences of the invention

XX SQ Sequence 344 BP; 146 A; 53 C; 85 G; 59 T; 0 U; 1 Other;

Query Match 84.7%; Score 14.4; DB 4; Length 344;  
Best Local Similarity 93.8%; Pred. No. 1.4e+03;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 CCCACGCTTGGCCGAG 17  
|||||  
Db 252 CCCACGCTTGGCCGTAG 237

RESULT 63

AAH09788/c

ID AAH09788 standard; cDNA; 546 BP.

XX AC

XX AC

XX AC

XX 26-JUN-2001 (first entry)

XX DT

XX DE Human cDNA clone (3'-primer) SEQ ID NO:6623.

XX DE

XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX KW

XX OS Homo sapiens.

XX OS

XX EN EP1074617-A2.

XX EN

XX PD 07-FEB-2001.

XX PD

XX PF 28-JUL-2000; 2000EP-00116126.

XX PF

XX PR 29-JUL-1999; 99JP-00248036.

XX PR

XX PR 27-AUG-1999; 99JP-00300253.

XX PR

XX PR 11-JAN-2000; 2000JP-00118776.

XX PR

XX PR 02-MAY-2000; 2000JP-00183767.

XX PR

XX PR 09-JUN-2000; 2000JP-00241899.

XX PR

XX PA (HELI-) HELIX RES INST.

XX PA

XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX PI WPI; 2001-318749/34.

XX DR

XX DR

XX PT

XX PT

XX PT

XX PT

XX PT

XX PT

XX PT

XX PT

XX PT

XX PT

XX PT

XX PT

XX PT

XX PT

XX PT

XX PT

XX PT

XX PT

XX PT

XX PT

XX PT

XX PT

CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893  
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent  
CC oligonucleotides, all of which are used in the exemplification of the  
CC present invention

XX SQ Sequence 546 BP; 119 A; 130 C; 124 G; 165 T; 0 U; 8 Other;

Query Match 84.7%; Score 14.4; DB 4; Length 546;  
Best Local Similarity 93.8%; Pred. No. 1.4e+03;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCCGA 16  
|||||  
Db 505 GCCCAGCTTGGCCGA 490

RESULT 64

ABV54022

ID ABV54022 standard; cDNA; 570 BP.

XX AC

XX AC

XX AC

XX 17-SEP-2002 (first entry)

XX DT

XX DE Human prostate expression marker cDNA 54013.

XX DE

XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

The invention relates to an isolated nucleic acid molecule (I) comprising  
a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
specification or its complement. (I) is useful for: (a) assessing whether  
a patient is afflicted with prostate cancer; (b) monitoring the  
progression of prostate cancer in a patient; (c) assessing the efficacy  
of a test compound to inhibit prostate cancer in a patient; (d) assessing  
the efficacy of a therapy for inhibiting prostate cancer in a patient;  
(e) selecting a composition for inhibiting prostate cancer in a patient;  
(f) assessing the prostate cell carcinogenic potential of a compound; (g)  
determining whether prostate cancer has metastasized in a patient; (h)  
assessing the aggressiveness or indolence of prostate cancer in a patient  
; (I) is also useful as a pharmacodynamic or pharmacogenomic marker

```
SQ Sequence 570 BP; 119 A; 167 C; 152 G; 129 T; 0 U; 3 Other;
Query Match      84.7%; Score 14.4; DB 5; Length 570;
Best Local Similarity 93.8%; Pred. No. 1.4e+03;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGA 16
    |||||
Db 140 GCCCAGCGTTGACCGA 155

RESULT 65
AAH08354
ID AAH08354 standard; cDNA; 680 BP.
XX
AC AAH08354;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA clone (5'-primer) SEQ ID NO:5189.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
FN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-00116126.
XX
PR 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX
PA (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
XX
PS Claim 1; SEQ ID NO 5189; 2537pp + Sequence Listing; English.
XX
XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises:
CC (a) an oligo-dT primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention

SQ Sequence 680 BP; 143 A; 194 C; 151 G; 184 T; 0 U; 8 Other;
Query Match      84.7%; Score 14.4; DB 4; Length 680;
Best Local Similarity 93.8%; Pred. No. 1.4e+03;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGA 16
    |||||
Db 633 GCCCAGCGTTGGCCA 648

RESULT 66
AAL55183/c
ID AAL55183 standard; DNA; 992 BP.
XX
AC AAL55183;
XX
DT 01-MAY-2003 (first entry)
XX
DE Verticillium dahliae defoliating plant DNA, SEQ ID No 2.
XX
KW Defoliating; non-defoliating; pathogen; Verticillium dahliae; nested PCR; olive; cotton; epidemiology; plant; ds.
XX
OS Verticillium dahliae.
XX
FN WO200297131-A1.
XX
PD 05-DEC-2002.
XX
PF 29-MAY-2002; 2002WO-ES000258.
XX
PR 29-MAY-2001; 2001ES-00001240.
XX
PA (NEWB-) NEWBIOTECHNIC SA.
XX
PA (UYCO-) UNIV CORDOBA.
XX
XX Mercado Blanco J, Rodriguez Jurado D, Perez Artes E;
PI Jimenez Diaz RM;
XX
DR WPI; 2003-140494/13.
XX
XX Identifying D and ND pathotypes of Verticillium dahliae, useful for early diagnosis of infection in olives and cotton, by nested polymerase chain reaction.
XX
PS Disclosure; Fig 2; 68pp; Spanish.
XX
XX The invention relates to a novel method for identifying defoliating and/or non-defoliating pathogenic types of Verticillium dahliae in a sample of plant material by use of a nested polymerase chain reaction. The method is used to differentiate between the pathotypes of V. dahliae of olives and cotton, e.g. to establish certification of disease-free plants, and also to study the epidemiology and control of V. dahliae. This polynucleotide sequence represents the DNA of a Verticillium dahliae defoliating plant of the invention
XX
SQ Sequence 992 BP; 212 A; 229 C; 306 G; 245 T; 0 U; 0 Other;
Query Match      84.7%; Score 14.4; DB 8; Length 992;
Best Local Similarity 93.8%; Pred. No. 1.4e+03;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGA 16
    |||||
Db 432 GCCCAGCGTTGGCCA 417

RESULT 67
AAV63901/c
ID AAV63901 standard; DNA; 1520 BP.
XX
```

AC AAV63901;  
 DT 20-JAN-1999 (first entry)  
 XX Helicobacter pylori DNA insert full restriction map of plasmid pILL405.  
 DE Helicobacter pylori; amIE; aliphatic amidase; gastritis; dyspepsia;  
 XX chronic gastroduodenal disorder; peptic ulcer; porcine gastric ulcer;  
 KW detection; infection; ss.  
 XX Helicobacter pylori.  
 OS WO9844094-A2.  
 XX 08-OCT-1998.  
 XX 27-MAR-1999; 98WO-EP001824.  
 PF 28-MAR-1997; 97US-0041745P.  
 XX (INSP ) INST PASTEUR.  
 XX (INRM ) INST NAT SANTE & RECH MEDICALE.  
 PI De Reuse H, Skouloubris S, Labigne A;  
 XX WPI; 1998-557106/47.  
 XX New Helicobacter aliphatic amidase AmIE polypeptides and their encoding  
 PT sequences - used in diagnosis, treatment and prevention of Helicobacter  
 PT sp. infections in humans and animals.  
 XX Claim 2; Fig 4b; 51pp; English.  
 PS  
 XX The present sequence represents the Helicobacter pylori DNA insert full  
 CC restriction map of plasmid pILL405. The present invention describes H.  
 CC pylori aliphatic amidase AmIE. AmIE polypeptides and antibodies form  
 CC immunogenic, pharmaceutical and therapeutic compositions which can be  
 CC used in methods for detecting, treating or preventing Helicobacter sp.  
 CC (particularly H. pylori and H. heilmannii) infections, especially chronic  
 CC gastroduodenal disorders like gastritis, dyspepsia and peptic ulcers in  
 CC man, and porcine gastric ulcers in pigs. They are administered in  
 CC immunologically/pharmacologically effective amounts by an oral,  
 CC intradermal, intramuscular, intravenous or mucosal route to a patient. A  
 CC detection kit for a Helicobacter infection comprises any AmIE  
 CC polynucleotides with any of the vectors given in the present invention.  
 CC The polypeptides are used to in a process to degrade acrylamide,  
 CC acetamide, propionamide and isobutyramide. The polypeptides are also used  
 CC to screen for active substrates that inhibit Helicobacter sp. amidase  
 CC activity. The polynucleotides encoding the aliphatic amidase AmIE  
 CC polypeptides are the first to be characterized at the molecular level.  
 CC Methods involving such polypeptides are preferred to urease-based methods  
 CC because of the presence of urea positive bacteria in porcine  
 CC gastrointestinal tracts  
 XX  
 SQ Sequence 1520 BP; 461 A; 261 C; 385 G; 413 T; 0 U; 0 Other;  
 Query Match 84.7%; Score 14.4; DB 2; Length 1520;  
 Best Local Similarity 93.8%; Pred. No. 1.4e+03;  
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 CCCAGCTTGGCCGAG 17  
 DB 1454 CCCAGCTTGGCCGAG 1439  
 RESULT 68  
 ACD05884/c  
 ID ACD05884 standard; cDNA; 1531 BP.  
 XX AC ACD05884;  
 XX 06-AUG-2003 (first entry)  
 DT  
 XX

DE Novel human contig #58.  
 XX Human; angiogenesis; cytokine; cell proliferation; pluripotent;  
 KW cell differentiation; totipotent; stem cell; transplantation; bio-sensor;  
 KW neuroepithelial cell; autoimmune disease; neural cell; genetic disorder;  
 KW nerve; brain tissue; central nervous system disease;  
 KW peripheral nervous system disease; neuropathy; haematopoiesis; bone;  
 KW myeloid disorder; lymphoid cell disorder; platelet disorder; tendon;  
 KW regeneration; cartilage; tendon; ligament; nerve tissue growth;  
 KW tissue repair; wound healing; burn; ulcer; osteoporosis; cancer;  
 KW osteoarthritis; bone degenerative disorder; periodontal disease;  
 KW gut protection; lung fibrosis; liver fibrosis; reperfusion injury;  
 KW immune deficiency; infection; autoimmune disorder; allergic reaction;  
 KW thrombolysis; thrombosis; coagulation disorder; hereditary disorder;  
 KW biorhythm; circadian cycle; fertility; metabolism; catabolism; anabolism;  
 KW neurotropic; neuroprotective; antiparkinsonian; anticonvulsant;  
 KW haemostatic; vulnery; antiulcer; osteopathic; antiarthritic;  
 KW vasotropic; immunostimulant; antibacterial; fungicide; immunosuppressive;  
 KW antirheumatic; antidiabetic; antiasthmatic; cytostatic; virucide;  
 KW expressed sequence tag; EST; ss.  
 XX Homo sapiens.  
 OS WO2003023013-A2.  
 XX 20-MAR-2003.  
 PD 13-SEP-2002; 2002WO-US029001.  
 PF 13-SEP-2001; 2001US-0322511P.  
 PR 12-SEP-2002; 2002US-00243552.  
 XX (HYSE-) HYSEQ INC.  
 PA Tang YT, Yang Y, Wang Z, Weng G, Ma Y;  
 XX WPI; 2003-313249/30.  
 PI P-PSDB; ABO00807.  
 DR Novel nucleic acids and polypeptides for diagnosis, treatment of central  
 PT and peripheral nervous system diseases and neuropathies, such as  
 PT Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 PT lateral sclerosis.  
 XX Example 2; SEQ ID NO 730; 300pp; English.  
 PS The present invention relates to the isolation of novel human  
 CC polynucleotide sequences and their encoding polypeptides. The novel  
 CC polypeptides exhibit activities relating to angiogenesis, cytokine, cell  
 CC proliferation, cell differentiation, antiinflammatory, and stem cell  
 CC growth factor activities. The polypeptides are involved in the  
 CC proliferation, differentiation and survival of pluripotent and totipotent  
 CC stem cells, and are useful for re-engineering damaged or diseased  
 CC tissues, transplantation, manufacture of bio-pharmaceuticals and  
 CC development of bio-sensors. The polypeptides can be used to manipulate  
 CC stem cells in culture to give rise to neuroepithelial cells that can be  
 CC used to augment or replace cells damaged by illness, autoimmune disease,  
 CC accidental damage or genetic disorders. The polypeptides induce the  
 CC proliferation of neural cells and regeneration of nerve and brain tissue  
 CC and are useful for the treatment of central and peripheral nervous system  
 CC diseases and neuropathies, such as Alzheimer's, Parkinson's disease,  
 CC Huntington's disease, amyotrophic lateral sclerosis (ALS). The  
 CC polypeptides are also involved in chemotactic or chemokinetic activity,  
 CC regulation of haematopoiesis and are useful for treating myeloid or  
 CC lymphoid cell disorders, platelet disorders such as thrombocytopaenia and  
 CC for regeneration of bone, cartilage, tendon, ligament and/or nerve tissue  
 CC growth, in tissue repair, healing of burns, incisions, ulcers, for  
 CC treating osteoporosis, osteoarthritis, bone degenerative disorders, and  
 CC periodontal disease. The polypeptides are also useful for gut protection  
 CC or regeneration and treatment of lung or liver fibrosis, reperfusion  
 CC injury in various tissues, various immune deficiencies and disorders  
 CC including severe combined immunodeficiency (SCID), bacterial or fungal  
 CC infections, autoimmune disorders (e.g. multiple sclerosis, rheumatoid

CC arthritis, diabetes mellitus, myasthenia gravis), allergic reactions and  
 CC conditions, such as asthma or other respiratory problems. The  
 CC polypeptides are involved in thrombolysis or thrombosis and are useful in  
 CC treatment of various coagulation disorders (including hereditary  
 CC disorders such as haemophilia) or to enhance coagulation and other  
 CC homeostatic events in treating wounds resulting from trauma, surgery or  
 CC other causes. The polypeptides exhibit immune stimulating or immune  
 CC suppressing activity, and are useful for treating autoimmune diseases or  
 CC cancer. They also inhibit the growth, infection or function of infectious  
 CC agents such as bacteria, fungi, viruses, effect biorhythms or circadian  
 CC cycles of rhythms, fertility of male or female subjects, metabolism,  
 CC catabolism, and anabolism. ACD05827-ACD06027 represent novel contigs  
 CC assembled using expressed sequence tag (EST) sequences as seeds. Note:  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 1531 BP; 407 A; 409 C; 451 G; 264 T; 0 U; 0 Other;

Query Match 84.7%; Score 14.4; DB 8; Length 1531;  
 Best Local Similarity 93.8%; Pred. No. 1.4e+03;  
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCCGA 16  
 Db 120 GCCCATGCTTGGCCGA 105  
 ||||| ||||| |||||

RESULT 69  
 AAV63900/C  
 ID AAV63900 standard; DNA; 1675 BP.  
 XX AC AAV63900;  
 XX  
 XX 20-JAN-1999 (first entry)  
 DT  
 XX Helicobacter pylori DNA insert sequence of plasmid pILL405.  
 DE  
 XX Helicobacter pylori; amIE; aliphatic amidase; gastritis; dyspepsia;  
 KW chronic gastroduodenal disorder; peptic ulcer; porcine gastric ulcer;  
 KW detection; infection; ss.  
 KW  
 XX Helicobacter pylori.  
 OS

XX WO9844094-A2.  
 XX  
 XX 08-OCT-1998.  
 XX  
 XX 27-MAR-1998; 98WO-EP001824.  
 XX  
 XX 28-MAR-1997; 97US-0041745P.  
 XX  
 XX (INSP ) INST PASTEUR.  
 XX (INRM ) INST NAT SANTE & RECH MEDICALE.  
 XX  
 XX De Reuse H, Skouloubris S, Labigne A;  
 XX WPI; 1998-557106/47.  
 XX  
 XX New Helicobacter aliphatic amidase AmIE polypeptides and their encoding  
 PT sequences - used in diagnosis, treatment and prevention of Helicobacter  
 PT sp. infections in humans and animals.  
 XX  
 XX Claim 2; Fig 4a; 51pp; English.

XX The present sequence represents the Helicobacter pylori DNA insert  
 CC sequence of plasmid pILL405. The present invention describes H. pylori  
 CC aliphatic amidase AmIE. AmIE polypeptides and antibodies form  
 CC immunogenic, pharmaceutical and therapeutic compositions which can be  
 CC used in methods for detecting, treating or preventing Helicobacter sp.  
 CC (particularly H. pylori and H. heilmannii) infections, especially chronic  
 CC gastroduodenal disorders like gastritis, dyspepsia and peptic ulcers in  
 CC man, and porcine gastric ulcers in pigs. They are administered in

CC immunologically/pharmacaceutically effective amounts by an oral,  
 CC intradermal, intramuscular, intravenous or mucosal route to a patient. A  
 CC detection kit for a Helicobacter infection comprises any AmIE  
 CC polynucleotides with any of the vectors given in the present invention.  
 CC The polypeptides are used in a process to degrade acrylamide,  
 CC acetamide, propionamide and isobutyramide. The polypeptides are also used  
 CC to screen for active substrates that inhibit Helicobacter sp. amidase  
 CC activity. The polynucleotides encoding the aliphatic amidase AmIE  
 CC polypeptides are the first to be characterized at the molecular level.  
 CC Methods involving such polypeptides are preferred to urease-based methods  
 CC because of the presence of urea positive bacteria in porcine  
 CC gastrointestinal tracts  
 XX  
 SQ Sequence 1675 BP; 503 A; 298 C; 424 G; 450 T; 0 U; 0 Other;

Query Match 84.7%; Score 14.4; DB 2; Length 1675;  
 Best Local Similarity 93.8%; Pred. No. 1.4e+03;  
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 CCCACGCTTGGCCGAG 17  
 Db 1534 CCCACGCTTGGCCGAG 1519  
 ||||| ||||| |||||

RESULT 70  
 ABZ35358/c  
 ID ABZ35358 standard; cDNA; 1889 BP.

XX AC ABZ35358;

XX 05-FEB-2003 (first entry)

XX Human gene expression profile polynucleotide SEQ ID NO 469.

XX Human; artery; endothelium; umbilical; vein; aorta; pulmonary artery;  
 KW bronchial epithelium; prostate; muscle; lung fibroblast; osteoblast;  
 KW tumour; microarray; genome mapping; antibiotic; antiviral; antifungal;  
 KW gene expression; gene; ss.

XX Homo sapiens.

XX WO200274979-A2.

XX 26-SEP-2002.

XX 20-MAR-2002; 2002WO-US008456.

XX 20-MAR-2001; 2001US-0276947P.

XX (ORTH ) ORTHO CLINICAL DIAGNOSTICS INC.

XX Wan J, Wang Y;

XX WPI; 2002-740862/80.

XX New gene expression profile generated from primary, endothelial,  
 PT epithelial, and muscle cell types, useful for identifying disease  
 PT pathologies involving alterations of gene expression, e.g. cancer.

XX Disclosure; Page 617-618; 850pp; English.

XX The invention relates to a gene expression profile comprising one or more  
 CC genes (ABZ34889-ABZ35692) and generated from a cell type. The cell type  
 CC is a coronary artery endothelium, umbilical artery or vein endothelium,  
 CC aortic endothelium, dermal microvascular endothelium, pulmonary artery  
 CC endothelium, myometrium microvascular endothelium, keratinocyte  
 CC epithelium, bronchial epithelium, mammary epithelium, prostate  
 CC epithelium, renal cortical epithelium, renal proximal tubule epithelium,  
 CC small airway epithelium, renal epithelium, umbilical artery smooth  
 CC muscle, neonatal dermal fibroblast, pulmonary artery smooth muscle,  
 CC dermal fibroblast, neural progenitor cells, skeletal muscle, astrocytes,  
 CC aortic smooth muscle, mesangial cells, coronary artery smooth muscle,  
 CC bronchial smooth muscle, uterine smooth muscle, lung fibroblast,



CC osteoblasts or prostate stromal cell. The gene expression profile is used  
CC for determining the level of RNA expression for a sample, determining the  
CC phenotype of a cell and distinguishing cell types. The gene or a protein  
CC expression profile is useful in identifying disease pathologies involving  
CC alterations of gene expression. The assessment of expression profiles may  
CC provide meaningful information with respect to tumour type and stage,  
CC treatment methods, and prognosis. The gene or protein expression profile  
CC may also be used for creating microarrays. The microarray is useful for  
CC genetic and physical mapping of genomes, DNA sequencing, genetic or  
CC medical diagnosis, genotyping of organisms, confirming cell or tissue  
CC antifungal agents  
CC  
SQ Sequence 1889 BP; 476 A; 464 C; 494 G; 455 T; 0 U; 0 Other;

Query Match 84.7%; Score 14.4; DB 6; Length 1889;  
Best Local Similarity 93.8%; Pred. No. 1.4e+03;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 CCCACGCTTGCCGAG 17  
D5 18 CCCACGCTTGCCGAG 3

RESULT 71  
ACN40151/c  
ID ACN40151 standard; cDNA; 3681 BP.  
XX  
AC ACN40151;  
XX  
DT 18-NOV-2004 (first entry)  
XX  
DE Tumour-associated antigenic target (TAT) cDNA DNA326301, SEQ ID NO:4752.  
XX  
KW Tumour-associated antigenic target; TAT; human; overexpression; cancer;  
KW tumour; diagnosis; cell proliferative disorder; breast cancer;  
KW colorectal cancer; lung cancer; ovarian cancer; liver cancer;  
KW central nervous system cancer; bladder cancer; pancreatic cancer;  
KW cervical cancer; melanoma; leukaemia; hybridisation probe;  
KW chromosome identification; chromosome mapping; gene mapping;  
KW gene therapy; cytostatic; gene; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO2004030615-A2.  
XX  
PD 15-APR-2004.  
XX  
PF 29-SEP-2003; 2003WO-US028547.  
XX  
PR 02-OCT-2002; 2002US-0414971P.  
XX  
PA (GETH) GENENTECH INC.  
XX  
PI Wu TD, Zhang Z, Zhou Y;  
XX  
DR WPI; 2004-347921/32.  
DR P-PSDB; ABM81845.  
XX  
PT New tumor-associated antigenic target polypeptides and nucleic acids,  
PT useful in preparing a medicament for treating or detecting a  
PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or  
PT prostate cancer or tumor.  
XX  
PS Claim 1; SEQ ID NO 4752; 7273pp; English.  
XX  
CC The invention relates to human tumour-associated antigenic target (TAT)  
CC polypeptides, and their related nucleic acids. The TAT polypeptides are  
CC overexpressed in cancer tissues compared to normal tissues, and may thus  
CC serve as effective targets for the diagnosis and treatment of cancer in  
CC mammals. The invention also relates to nucleic acid and polypeptide  
CC sequences at least 80% identical to the TAT nucleic acids and  
CC polypeptides; expression vectors and host cells comprising a TAT nucleic

CC acid; an antibody specific for a TAT polypeptide; a peptide or organic  
CC molecule which binds to a TAT polypeptide; fusion proteins comprising a  
CC TAT polypeptide; and methods and compositions for the treatment or  
CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,  
CC antibodies, antagonists, binding molecules and compositions are useful  
CC for diagnosing or treating a cell proliferative disorder associated with  
CC increased TAT expression, particularly cancers such as breast cancer,  
CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder  
CC cancer, pancreatic cancer, cervical cancer, cancers of the central  
CC nervous system, melanoma and leukaemia. TAT nucleic acids may further be  
CC used as hybridisation probes, in chromosome and gene mapping, in  
CC chromosome identification and in gene therapy. The present sequence  
CC represents a TAT nucleic acid of the invention  
XX  
SQ Sequence 3681 BP; 1095 A; 903 C; 968 G; 715 T; 0 U; 0 Other;

Query Match 84.7%; Score 14.4; DB 13; Length 3681;  
Best Local Similarity 93.8%; Pred. No. 1.4e+03;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCCAGCTTGCCGA 16  
D5 149 GCCCAGCTTGCCGA 134

RESULT 72  
ABK89239/c  
ID ABK89239 standard; DNA; 13608 BP.  
XX  
AC ABK89239;  
XX  
DT 21-OCT-2002 (first entry)  
XX  
DE Human Na/Cl-dependent organic solute transporter gene.  
XX  
KW Human; ds; gene; Na/Cl-dependent organic solute cotransporter;  
KW chromosome 5; neuron tumour; single nucleotide polymorphism; SNP.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT variation replace(26,C)  
FT /\*tag= a  
FT /\*standard\_name= "Single nucleotide polymorphism"  
FT variation replace(36,R)  
FT /\*tag= b  
FT /\*standard\_name= "Single nucleotide polymorphism"  
FT variation replace(131,R)  
FT /\*tag= c  
FT /\*standard\_name= "Single nucleotide polymorphism"  
FT variation replace(1330,C)  
FT /\*tag= d  
FT /\*standard\_name= "Single nucleotide polymorphism"  
FT variation replace(1603,C)  
FT /\*tag= e  
FT /\*standard\_name= "Single nucleotide polymorphism"  
FT variation replace(1773,T)  
FT /\*tag= f  
FT /\*standard\_name= "Single nucleotide polymorphism"  
FT variation replace(1885,T)  
FT /\*tag= g  
FT /\*standard\_name= "Single nucleotide polymorphism"  
FT variation replace(1897,T)  
FT /\*tag= h  
FT /\*standard\_name= "Single nucleotide polymorphism"  
FT variation replace(2010,G)  
FT /\*tag= i  
FT /\*standard\_name= "Single nucleotide polymorphism"  
FT variation replace(2260,G)  
FT /\*tag= j  
FT /\*standard\_name= "Single nucleotide polymorphism"  
FT variation replace(2314,T)  
FT /\*tag= k

FT variation /standard\_name="Single nucleotide polymorphism"  
 FT replace(2650,C)  
 FT /\*tag= l  
 FT variation /standard\_name="Single nucleotide polymorphism"  
 FT replace(4944,A)  
 FT /\*tag= m  
 FT variation /standard\_name="Single nucleotide polymorphism"  
 FT replace(5230,T)  
 FT /\*tag= n  
 FT variation /standard\_name="Single nucleotide polymorphism"  
 FT replace(5481,G)  
 FT /\*tag= o  
 FT variation /standard\_name="Single nucleotide polymorphism"  
 FT replace(5765,T)  
 FT /\*tag= p  
 FT variation /standard\_name="Single nucleotide polymorphism"  
 FT replace(6097,A)  
 FT /\*tag= q  
 FT variation /standard\_name="Single nucleotide polymorphism"  
 FT replace(7194,V)  
 FT /\*tag= r  
 FT variation /standard\_name="Single nucleotide polymorphism"  
 FT replace(1117,G)  
 FT /\*tag= s  
 FT variation /standard\_name="Single nucleotide polymorphism"  
 FT replace(11398,T)  
 FT /\*tag= t  
 FT variation /standard\_name="Single nucleotide polymorphism"  
 FT replace(12464,A)  
 FT /\*tag= u  
 FT variation /standard\_name="Single nucleotide polymorphism"  
 FT replace(12736,C)  
 FT /\*tag= v  
 FT /standard\_name="Single nucleotide polymorphism"

XX WO200253741-A2.

XX 11-JUL-2002.

XX 02-JAN-2002; 2002WO-US000111.

XX 03-JAN-2001; 2001US-00752821.

PR 22-MAY-2001; 2001US-00861846.

XX (PEKE ) PE CORP NY.

XX Gugler K, Brandon RC, Di Francesco V, Beasley EM;

XX WPT; 2002-583618/62.

XX Novel isolated human peptide useful for treating disorders characterized by the absence of, inappropriate or unwanted expression of the receptor protein, and as immunogens to raise antibodies.

PS Claim 4; Fig 3; 88pp; English.

XX The invention relates to an isolated human peptide (S1) comprising an Na/C1-dependent organic solute cotransporter family member, an allelic variant or orthologue of (S1) encoded by a polynucleotide hybridising under stringent conditions to the opposite strand of a sequence appearing as ABK89238 and ABK9239, or a fragment of (S1) with at least 10 contiguous amino acids. A pharmaceutical composition comprising a modulator of S1 is useful for treating a disease or condition mediated by a human protease. S1 and its encoding nucleic acid are useful as models for the development of human therapeutics, for identifying therapeutic proteins, as targets for development of human therapeutic agents. S1 is useful to raise antibodies or to elicit another immune response, as a reagent in assays designed to quantitatively determine levels of the protein in biological fluids, as markers for tissues in which the corresponding protein is preferentially expressed, in drug screening assays, in cell-based or cell-free systems, to identify compounds that modulate cotransporter activity of the protein in its natural state, or an altered form that causes the specific disease or pathology associated

CC with the receptor (e.g. neuron tumour), as bait proteins in a two-hybrid or three-hybrid assay, to provide a target for diagnosing a disease or predisposition to disease mediated by the peptide, in pharmacogenomic analysis, and for treating a disorder characterised by an absence of, in appropriate or unwanted expression of S1. An anti-S1 antibody is useful for isolating and purifying S1, to detect S1 in cells or tissues, in situ, in vitro or in a cell lysate or supernatant, to assess abnormal tissue distribution or abnormal expression during development or progression of a biological condition, and for inhibiting a protein function. S1 nucleic acids are useful as probes, primers, and chemical intermediates, in biological assays, for constructing recombinant vectors, host cells and transgenic animals, for expressing antigenic portions of the S1 proteins, or for designing ribozymes. A vector comprising the S1 nucleic acid is useful for producing an S1 protein or peptide, for conducting cell-based assays involving the protein or its fragment, for identifying protein mutants whose functions are affected, and to produce non-human transgenic animals. The gene for the cotransporter is located on chromosome 5. The present sequence genomic DNA for the cotransporter protein

XX SQ Sequence 13608 BP; 2678 A; 4019 C; 3850 G; 2869 T; 0 U; 192 Other;

Query Match 84.7%; Score 14.4; DB 6; Length 13608;

Best Local Similarity 93.8%; Pred. No. 1.5e+03;

Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 CCCACGCTTGCCGAG 17

Db 7742 CCCACGCTTGCTGAG 7727

RESULT 73

AED48654  
 ID AED48654 standard; DNA; 48169 BP.

XX AC AED48654;

XX 15-DEC-2005 (first entry)

DE Streptomyces rimosus paromomycin biosynthetic gene cluster DNA.

XX ds; gene; antibiotic; virucide; aminoglycoside; bacterial infection;  
 KW viral infection; paromomycin; parE; parS; parC; parM; parT; parU; parQ;  
 KW parP; parR; parX; parF; parD; parU; parB; parA; parI; parR; parY;  
 KW parG; parH; P450 heme thiolate protein; cytochrome P450 hydroxylase;  
 KW non-ribosomal peptide synthetase; polyketide cyclase/reductase;  
 KW amidinotransferase; DNA-binding protein;  
 KW 3-amino-2,3-dideoxy-scyllo-inositol-1-dehydrogenase;  
 KW bifunctional L-glutamine:ketocyclitol aminotransferase I and II;  
 KW 2-deoxy-scyllo-inosose synthase;  
 KW UDP-Nac-glucosamine:2-deoxystreptamine-4-glycosyltransferase;  
 KW ABC transporter, ATP-binding protein; paromomycin 6'-dehydrogenase;  
 KW Fe-S oxidoreductase; phosphatase; glycosyltransferase;  
 KW 2-N-acetylhexosamine deacetylase; paromomycin 6'-aminotransferase;  
 KW extracellular aminoglycoside phosphate phosphatase;  
 KW aminoglycoside 3'-phosphotransferase; regulatory protein;  
 KW NDP-hexose synthase; chitinase;  
 KW nucleoside diphosphate-sugar oxidoreductase;  
 KW tetR family transcriptional regulator; transcriptional regulator;  
 KW oxidoreductase.

XX Streptomyces rimosus.

XX WO2005095591-A2.

XX 13-OCT-2005.

XX 15-MAR-2005; 2005WO-DE000499.

XX 01-APR-2004; 2004DE-10017141.

XX (COMB-) COMBINATURE BIOPHARM AG.

PA (UYWU-) UNIV WUPPERTAL.

XX Vente A, Piepersberg W, Wehmeier U, Schmidt-Beissner H;  
PI Aboshanab KMA, Welzel K;  
XX WPI; 2005-734187/75.  
DR P-PSDB; AED48614, AED48615, AED48616, AED48617, AED48618, AED48619,  
DR AED48620, AED48621, AED48622, AED48623, AED48624, AED48625, AED48626,  
DR AED48627, AED48628, AED48629, AED48630, AED48631, AED48632, AED48633,  
DR AED48634, AED48635, AED48636, AED48637, AED48638, AED48639, AED48640,  
DR AED48641, AED48642, AED48643, AED48644, AED48645, AED48646, AED48647,  
DR AED48648, AED48649, AED48650, AED48651, AED48652, AED48653.  
DR DDBJ; AJ628955.  
XX Isolated proteins or peptides that are expression products of genes  
PT derived from microorganisms producing aminoglycoside antibiotics.  
PT  
XX Example 1; Page 185-199; 367pp; German.  
XX This invention describes novel isolated proteins or peptides that are  
CC expression products of genes derived from aminoglycoside-producing  
CC microorganisms. The proteins or peptides are functional for a step in the  
CC biosynthesis of aminoglycoside antibiotics, especially butirosin,  
CC gentamicin, neomycin, fortimicin, tobramycin, apramycin, paromomycin,  
CC hygromycin B, ribostamycin, sisomicin, sisomicin, kanamycin, nebramycin,  
CC sel domycin, destomycin, istamycin, sannamycin, dactimicin, sporaricin,  
CC bluensomycin, ashimycin, lividomycin and spectinomycin. The invention  
CC also describes a) isolated nucleic acids encoding the proteins or  
CC peptides which are under the control of a regulatory sequence, especially  
CC a promoter; b) transformation vectors containing the nucleic acids; c)  
CC cells containing the nucleic acids and/or the proteins or peptides; d) a  
CC method of producing the cells by transforming cells with vectors as above  
CC ; e) aminoglycosides produced by culturing the cells; f) producing  
CC aminoglycosides by culturing cells as above and isolating aminoglycosides  
CC from the cells and/or culture supernatant and g) pharmaceutical  
CC compositions containing the aminoglycosides. The aminoglycosides are  
CC useful for treating bacterial or viral infections, especially infections  
CC with Gram-positive or -negative bacteria, e.g. *Escherichia coli* or  
CC *Klebsiella*. This sequence encodes proteins (AED48605-AED48653) from  
CC *Streptomyces rimosus* (DDBJ reference AJ628955) involved in the synthesis  
CC of paromomycin.  
XX  
SQ Sequence 48169 BP; 6567 A; 17743 C; 17423 G; 6436 T; 0 U; 0 Other;  
Query Match 84.7%; Score 14.4; DB 14; Length 48169;  
Best Local Similarity 93.8%; Pred. No. 1.5e+03;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GCCCAGCTTGCCGA 16  
DB 45829 GCCCAGCTTGCTGA 45844  
RESULT 74  
AAH51594/c  
ID AAH51594 standard; DNA; 49312 BP.  
XX  
XX AC AAH51594;  
XX  
XX 29-AUG-2001 (first entry)  
XX  
XX Human genomic sequence of the MGST-II gene SEQ ID 485.  
XX Human; biallelic marker; single nucleotide polymorphism; SNP; MGSTII;  
KW microsomal glutathione S-transferase II; malate decarboxylase enzyme;  
KW DMEI; MEI; cytochrome P450; glutathione reductase; GSHR; GSHS; GGT5;  
KW flavin-containing monooxygenase; FMO; gamma-glutamyltransferase 5;  
KW dipeptidase; DP; glucose 6-phosphate dehydrogenase; G6PDH; haplotype;  
KW phosphogluconate dehydrogenase; PGDH; drug metabolism; phenotype;  
KW uridine diphosphate glucuronosyl transferase; UGT2; asthma; hepatotoxicity;  
KW zileuton; ds.  
XX  
XX OS Homo sapiens.  
XX

PN WO200058508-A2.  
XX  
PD 05-OCT-2000.  
XX  
XX 24-MAR-2000; 2000WO-IB000403.  
PF  
XX 25-MAR-1999; 99US-0126269P.  
PR 30-APR-1999; 99US-0131961P.  
XX (GEST ) GENSET.  
PA  
XX Blumenfeld M, Bougueleret L, Chumakov I, Cohen-Akenine A;  
PI WPI; 2000-638353/61.  
XX Polynucleotides comprising sequences from malate decarboxylase enzyme-  
DR related biallelic markers used for genotyping.  
PT  
XX Claim 88; Page 656-669; 673pp; English.  
PS  
XX Sequences AAH51110-AAH51593 represent human DNA fragments which contain  
CC biallelic markers. The sequences are related to various human genes  
CC including microsomal glutathione S-transferase II (MGSTII), malate  
CC decarboxylase enzyme (DMEI/MEI), cytochrome P450, glutathione  
CC reductase/synthase (GSHR/GSHS), flavin-containing monooxygenases (FMO),  
CC gamma-glutamyltransferase 5 (GGT5), dipeptidase (DP), glucose 6-phosphate  
CC dehydrogenase (G6PDH), phosphogluconate dehydrogenase (PGDH), and uridine  
CC diphosphate glucuronosyl transferases (UGT2). Each of these sequences  
CC contains a biallelic marker/polymorphism, which is represented in the  
CC sequence as a degenerate/undefined base. The genes to which the biallelic  
CC marker containing sequences are related are involved in drug metabolism.  
CC Sequences AAH51594 - AAH51598 represent the genomic sequence of the  
CC MGSTII gene and four alternative MGSTII cDNA sequences. AAB62905-AAB62906  
CC are MGSTII gene products. PCR primers AAH51599 and AAH51600 are used in  
CC an example for the amplification of human genomic DNA fragments. The  
CC invention includes a method of genotyping comprising determining the  
CC identity of a nucleotide at a DME- or MGSTII-related biallelic marker in  
CC a biological sample. The method is used to determine the frequency in  
CC a population of an allele of a DME- or MGST-II related biallelic marker and  
CC to select an individual for inclusion in a clinical trial of a drug  
CC treatment. The method is also used to detect association between allele  
CC and phenotype, and to detect association between haplotype and phenotype.  
CC The polynucleotides are used, in hybridization assays, sequencing assays  
CC or allele specific amplification assays. The method can be used to  
CC determine whether an individual suffers or is at risk of developing  
CC asthma or is at risk of developing hepatotoxicity on treatment with  
CC zileuton  
XX  
SQ Sequence 49312 BP; 13125 A; 10209 C; 11088 G; 14881 T; 0 U; 9 Other;  
Query Match 84.7%; Score 14.4; DB 3; Length 49312;  
Best Local Similarity 93.8%; Pred. No. 1.5e+03;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 CCCACGCTTGCCGAG 17  
DB 12970 CCCACTCTTGCCGAG 12955  
RESULT 75  
ADN46845\_20  
Continuation (21 of 21) of ADN46845 from base 2000001 (Thermococcus kodakaraensis KOD1  
WP Sequence split into 21 fragments LOCUS ADN46845 Accession Adn46845  
WP Fragment Name Begin End  
WP ADN46845\_00 1 110000  
WP ADN46845\_01 100001 210000  
WP ADN46845\_02 200001 310000  
WP ADN46845\_03 300001 410000  
WP ADN46845\_04 400001 510000  
WP ADN46845\_05 500001 610000  
WP ADN46845\_06 600001 710000  
WP ADN46845\_07 700001 810000  
WP ADN46845\_08 800001 910000

WP ADN46845\_09 900001 1010000  
WP ADN46845\_10 1000001 1110000  
WP ADN46845\_11 1100001 1210000  
WP ADN46845\_12 1200001 1310000  
WP ADN46845\_13 1300001 1410000  
WP ADN46845\_14 1400001 1510000  
WP ADN46845\_15 1500001 1610000  
WP ADN46845\_16 1600001 1710000  
WP ADN46845\_17 1700001 1810000  
WP ADN46845\_18 1800001 1910000  
WP ADN46845\_19 1900001 2010000  
WP ADN46845\_20 2000001 2089378

Query Match 84.7%; Score 14.4; DB 12; Length 89378;  
Best Local Similarity 93.8%; Pred. No. 1.5e+03;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCACGCTGGCCGA 16  
|||||

Db 64028 GCCCACGCTGGCCGA 64043

## RESULT 76

ADN46123\_20

Continuation (21 of 21) of ADN46123 from base 2000001 (Thermococcus kodakaraensis KOD1  
WP Sequence split into 21 fragments LOCUS ADN46123 Accession Adn46123

WP Fragment Name Begin End  
WP ADN46123\_00 1 110000  
WP ADN46123\_01 100001 210000  
WP ADN46123\_02 200001 310000  
WP ADN46123\_03 300001 410000  
WP ADN46123\_04 400001 510000  
WP ADN46123\_05 500001 610000  
WP ADN46123\_06 600001 710000  
WP ADN46123\_07 700001 810000  
WP ADN46123\_08 800001 910000  
WP ADN46123\_09 900001 1010000  
WP ADN46123\_10 1000001 1110000  
WP ADN46123\_11 1100001 1210000  
WP ADN46123\_12 1200001 1310000  
WP ADN46123\_13 1300001 1410000  
WP ADN46123\_14 1400001 1510000  
WP ADN46123\_15 1500001 1610000  
WP ADN46123\_16 1600001 1710000  
WP ADN46123\_17 1700001 1810000  
WP ADN46123\_18 1800001 1910000  
WP ADN46123\_19 1900001 2010000  
WP ADN46123\_20 2000001 2089378

Query Match 84.7%; Score 14.4; DB 12; Length 89378;  
Best Local Similarity 93.8%; Pred. No. 1.5e+03;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCACGCTGGCCGA 16  
|||||

Db 64028 GCCCACGCTGGCCGA 64043

## RESULT 77

ADN46464\_20

Continuation (21 of 21) of ADN46464 from base 2000001 (Thermococcus kodakaraensis KOD1  
WP Sequence split into 21 fragments LOCUS ADN46464 Accession Adn46464

WP Fragment Name Begin End  
WP ADN46464\_00 1 110000  
WP ADN46464\_01 100001 210000  
WP ADN46464\_02 200001 310000  
WP ADN46464\_03 300001 410000  
WP ADN46464\_04 400001 510000  
WP ADN46464\_05 500001 610000  
WP ADN46464\_06 600001 710000  
WP ADN46464\_07 700001 810000  
WP ADN46464\_08 800001 910000  
WP ADN46464\_09 900001 1010000

WP ADN46464\_10 1000001 1110000  
WP ADN46464\_11 1100001 1210000  
WP ADN46464\_12 1200001 1310000  
WP ADN46464\_13 1300001 1410000  
WP ADN46464\_14 1400001 1510000  
WP ADN46464\_15 1500001 1610000  
WP ADN46464\_16 1600001 1710000  
WP ADN46464\_17 1700001 1810000  
WP ADN46464\_18 1800001 1910000  
WP ADN46464\_19 1900001 2010000  
WP ADN46464\_20 2000001 2089378

Query Match 84.7%; Score 14.4; DB 12; Length 89378;  
Best Local Similarity 93.8%; Pred. No. 1.5e+03;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCACGCTGGCCGA 16  
|||||

Db 64028 GCCCACGCTGGCCGA 64043

## RESULT 78

ADN47591\_00/c

WP Sequence split into 21 fragments LOCUS ADN47591 Accession Adn47591

WP Fragment Name Begin End  
WP ADN47591\_00 1 110000  
WP ADN47591\_01 100001 210000  
WP ADN47591\_02 200001 310000  
WP ADN47591\_03 300001 410000  
WP ADN47591\_04 400001 510000  
WP ADN47591\_05 500001 610000  
WP ADN47591\_06 600001 710000  
WP ADN47591\_07 700001 810000  
WP ADN47591\_08 800001 910000  
WP ADN47591\_09 900001 1010000  
WP ADN47591\_10 1000001 1110000  
WP ADN47591\_11 1100001 1210000  
WP ADN47591\_12 1200001 1310000  
WP ADN47591\_13 1300001 1410000  
WP ADN47591\_14 1400001 1510000  
WP ADN47591\_15 1500001 1610000  
WP ADN47591\_16 1600001 1710000  
WP ADN47591\_17 1700001 1810000  
WP ADN47591\_18 1800001 1910000  
WP ADN47591\_19 1900001 2010000  
WP ADN47591\_20 2000001 2089378  
ID ADN47591 standard; DNA; 2089378 BP.  
XX  
AC ADN47591;

XX  
DT 01-JUL-2004 (first entry)

XX  
DE Thermococcus kodakaraensis KOD1 genome DNA sequence SeqID1469.

XX  
KW gene disruption; gene targeting; marker gene; transformation;  
KW homologous recombination; hyperthermostable archaeobacterium; KOD1;  
KW gene structure; gene function; enzyme activity; medicine;  
KW forensic science; food; drug inspection; molecular biology; immunology;  
KW ds; gene.

XX  
OS Thermococcus kodakaraensis.

XX  
PN WO2004022736-A1.

XX  
PD 18-MAR-2004.

XX  
PF 29-AUG-2003; 2003WO-IB003597.

XX  
PR 30-AUG-2002; 2002JP-00319011.

XX  
PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.

XX  
PI Imanaka T, Atomi H;

```
XX WPI; 2004-257583/24.
XX
XX KW gene disruption; gene targeting; marker gene; transformation;
XX KW homologous recombination; hyperthermostable archaeobacterium; KOD1;
XX KW gene structure; gene function; enzyme activity; medicine;
XX KW forensic science; food; drug inspection; molecular biology; immunology;
XX KW ds; gene.
XX
XX OS Thermococcus kodakaraensis.
XX
XX PN WO2004022736-A1.
XX
XX PD 18-MAR-2004.
XX
XX PF 29-AUG-2003; 2003WO-IB003597.
XX
XX PR 30-AUG-2002; 2002JP-00319011.
XX
XX XX (NISC-) JAPAN SCI & TECHNOLOGY CORP.
XX
XX Imanaka T, Atomi H;
XX
XX WPI; 2004-257583/24.
XX
XX Method for disrupting targeted gene in genome of organism particularly
XX thermotable bacterium and with genome chips for analysis, applicable in
XX studying gene structure and functions.
XX
XX PS Claim 8; SEQ ID NO 1087; 598pp; Japanese.
XX
XX CC This invention relates to a novel method for targeting disruption of an
XX arbitrary gene in a genome of an organism which comprises providing the
XX whole sequential data of the genome of such organism, selecting at least
XX 1 arbitrary region in the sequence, providing a vector that contains a
XX sequence homologous with the selected region and a marker gene,
XX transformation, and homologous recombination. The genome is preferably
XX the genome of a hyperthermostable archaeobacterium, particularly
XX Thermococcus kodakaraensis KOD1. The method is for targeting the
XX disruption of a gene in the genome of an organism, which is applicable in
XX studying gene structure and functions as well as enzyme activities of
XX encoded proteins and useful in medicine, forensic science, food or drug
XX inspection, molecular biology and immunology. With this method, the
XX disruption of a gene at an arbitrary position in a genome can be achieved
XX efficiently and reliably. The present sequence is that of the genomic DNA
XX sequence of thermococcus kodakaraensis (KOD1) which was derived during
XX use of the method of the invention. Note: The sequence data for this
XX patent did not form part of the printed specification, but was obtained
XX in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX SQ Sequence 2089378 BP; 499376A; 542308C; 544081G; 503594T; 0U; 190ther;
XX
XX Query Match 84.7%; Score 14.4; DB 12; Length 110000;
XX Best Local Similarity 93.8%; Pred. No. 1.5e+03;
XX Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX Qy 1 GCCCAGCTTGGCCGA 16
XX |||||
XX Db 25350 GCCCAGCTGCGCGA 25335
XX
XX RESULT 79
XX ADN47209 00/c
XX WP Sequence split into 21 fragments LOCUS ADN47209 Accession Adn47209
XX WP Fragment Name Begin End
XX WP ADN47209_00 1 110000
XX WP ADN47209_01 100001 210000
XX WP ADN47209_02 200001 310000
XX WP ADN47209_03 300001 410000
XX WP ADN47209_04 400001 510000
XX WP ADN47209_05 500001 610000
XX WP ADN47209_06 600001 710000
XX WP ADN47209_07 700001 810000
XX WP ADN47209_08 800001 910000
XX WP ADN47209_09 900001 1010000
XX WP ADN47209_10 1000001 1110000
XX WP ADN47209_11 1100001 1210000
XX WP ADN47209_12 1200001 1310000
XX WP ADN47209_13 1300001 1410000
XX WP ADN47209_14 1400001 1510000
XX WP ADN47209_15 1500001 1610000
XX WP ADN47209_16 1600001 1710000
XX WP ADN47209_17 1700001 1810000
XX WP ADN47209_18 1800001 1910000
XX WP ADN47209_19 1900001 2010000
XX WP ADN47209_20 2000001 2089378
XX ID ADN47209 standard; DNA; 2089378 BP.
XX
XX AC ADN47209;
XX
XX DT 01-JUL-2004 (first entry)
XX
XX XX Thermococcus kodakaraensis KOD1 genome DNA sequence SeqID1087.
XX DE
XX XX
```

```
KW gene disruption; gene targeting; marker gene; transformation;
KW homologous recombination; hyperthermostable archaeobacterium; KOD1;
KW gene structure; gene function; enzyme activity; medicine;
KW forensic science; food; drug inspection; molecular biology; immunology;
KW ds; gene.
XX
XX OS Thermococcus kodakaraensis.
XX
XX PN WO2004022736-A1.
XX
XX PD 18-MAR-2004.
XX
XX PF 29-AUG-2003; 2003WO-IB003597.
XX
XX PR 30-AUG-2002; 2002JP-00319011.
XX
XX XX (NISC-) JAPAN SCI & TECHNOLOGY CORP.
XX
XX Imanaka T, Atomi H;
XX
XX WPI; 2004-257583/24.
XX
XX Method for disrupting targeted gene in genome of organism particularly
XX thermotable bacterium and with genome chips for analysis, applicable in
XX studying gene structure and functions.
XX
XX PS Claim 8; SEQ ID NO 1087; 598pp; Japanese.
XX
XX CC This invention relates to a novel method for targeting disruption of an
XX arbitrary gene in a genome of an organism which comprises providing the
XX whole sequential data of the genome of such organism, selecting at least
XX 1 arbitrary region in the sequence, providing a vector that contains a
XX sequence homologous with the selected region and a marker gene,
XX transformation, and homologous recombination. The genome is preferably
XX the genome of a hyperthermostable archaeobacterium, particularly
XX Thermococcus kodakaraensis KOD1. The method is for targeting the
XX disruption of a gene in the genome of an organism, which is applicable in
XX studying gene structure and functions as well as enzyme activities of
XX encoded proteins and useful in medicine, forensic science, food or drug
XX inspection, molecular biology and immunology. With this method, the
XX disruption of a gene at an arbitrary position in a genome can be achieved
XX efficiently and reliably. The present sequence is that of the genomic DNA
XX sequence of Thermococcus kodakaraensis (KOD1) which was derived during
XX use of the method of the invention. Note: The sequence data for this
XX patent did not form part of the printed specification, but was obtained
XX in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX SQ Sequence 2089378 BP; 499376A; 542308C; 544081G; 503594T; 0U; 190ther;
XX
XX Query Match 84.7%; Score 14.4; DB 12; Length 110000;
XX Best Local Similarity 93.8%; Pred. No. 1.5e+03;
XX Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX Qy 1 GCCCAGCTTGGCCGA 16
XX |||||
XX Db 25350 GCCCAGCTGCGCGA 25335
XX
XX RESULT 80
XX ADN47960 00/c
XX WP Sequence split into 21 fragments LOCUS ADN47960 Accession Adn47960
XX WP Fragment Name Begin End
XX WP ADN47960_00 1 110000
XX WP ADN47960_01 100001 210000
XX WP ADN47960_02 200001 310000
XX WP ADN47960_03 300001 410000
XX WP ADN47960_04 400001 510000
XX WP ADN47960_05 500001 610000
XX WP ADN47960_06 600001 710000
XX WP ADN47960_07 700001 810000
XX WP ADN47960_08 800001 910000
XX WP ADN47960_09 900001 1010000
```

WP ADN47960\_10 1000001 1110000  
WP ADN47960\_11 1100001 1210000  
WP ADN47960\_12 1200001 1310000  
WP ADN47960\_13 1300001 1410000  
WP ADN47960\_14 1400001 1510000  
WP ADN47960\_15 1500001 1610000  
WP ADN47960\_16 1600001 1710000  
WP ADN47960\_17 1700001 1810000  
WP ADN47960\_18 1800001 1910000  
WP ADN47960\_19 1900001 2010000  
WP ADN47960\_20 2000001 2093378  
ID ADN47960 standard; DNA; 2089378 BP.  
XX AC ADN47960;  
XX AC  
XX 01-JUL-2004 (first entry)  
XX DE Thermococcus kodakaraensis KOD1 genome DNA sequence SeqID1838.  
XX DE  
XX gene disruption; gene targeting; marker gene; transformation;  
KW homologous recombination; hyperthermostable archaeobacterium; KOD1;  
KW gene structure; gene function; enzyme activity; medicine;  
KW forensic science; food; drug inspection; molecular biology; immunology;  
KW ds; gene.  
XX OS Thermococcus kodakaraensis.  
XX OS  
XX WO2004022736-A1.  
XX PN  
XX 18-MAR-2004.  
XX PD  
XX 29-AUG-2003; 2003WO-IB003597.  
XX PF  
XX 30-AUG-2002; 2002JP-00319011.  
XX PR  
XX (NISC-) JAPAN SCI & TECHNOLOGY CORP.  
XX PA  
XX Imanaka T, Atomí H;  
XX PI  
XX WPI; 2004-257583/24.  
XX DR  
XX Method for disrupting targeted gene in genome of organism particularly  
PT thermostable bacterium and with genome chips for analysis, applicable in  
PT studying gene structure and functions.  
XX PT  
XX Example 1; SEQ ID NO 1838; 598pp; Japanese.  
XX PS  
XX This invention relates to a novel method for targeting disruption of an  
CC arbitrary gene in a genome of an organism which comprises providing the  
CC whole sequential data of the genome of such organism, selecting at least  
CC 1 arbitrary region in the sequence, providing a vector that contains a  
CC sequence homologous with the selected region and a marker gene,  
CC transformation, and homologous recombination. The genome is preferably  
CC the genome of a hyperthermostable archaeobacterium, particularly  
CC Thermococcus kodakaraensis KOD1. The method is for targeting the  
CC disruption of a gene in the genome of an organism, which is applicable in  
CC studying gene structure and functions as well as enzyme activities of  
CC encoded proteins and useful in medicine, forensic science, food or drug  
CC inspection, molecular biology and immunology. With this method, the  
CC disruption of a gene at an arbitrary position in a genome can be achieved  
CC efficiently and reliably. The present position is that of the genomic DNA  
CC sequence of Thermococcus kodakaraensis (KOD1) which was derived during  
CC use of the method of the invention. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX CC  
XX Sequence 2089378 BP; 499376A; 542308C; 503594T; 0U; 190ther;  
SQ  
Query Match 84.7%; Score 14.4; DB 12; Length 110000;  
Best Local Similarity 93.8%; Pred. No. 1.5e+03;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGGCGGA 16  
Db 25350 GCCACGCTGCGCGA 25335  
RESULT 81  
ABD32843  
ID ABD32843 standard; DNA; 326002 BP.  
XX AC ABD32843;  
XX AC  
XX 18-NOV-2004 (first entry)  
XX DT  
XX DE Human cancer-associated genomic DNA HD17-014.  
XX DE  
XX Human; ds; cancer-associated protein; gene; cytostatic; cancer;  
KW leukaemia; lymphoma; CAP.  
KW  
XX OS Homo sapiens.  
XX OS  
XX WO2004074320-A2.  
XX PN  
XX 02-SEP-2004.  
XX PD  
XX 17-FEB-2004; 2004WO-US004730.  
XX PF  
XX 14-FEB-2003; 2003US-00367094.  
XX PR  
XX 14-MAR-2003; 2003US-00388838.  
XX PR  
XX 15-APR-2003; 2003US-00417375.  
XX PR  
XX 13-JUN-2003; 2003US-00461862.  
XX PR  
XX 15-SEP-2003; 2003US-00663431.  
XX PR  
XX 15-DEC-2003; 2003US-00737318.  
XX PR  
XX (SAGR-) SAGRES DISCOVERY INC.  
XX PA  
XX Morris DW, Morris DW, Malandro MS;  
XX PI  
XX WPI; 2004-652914/63.  
XX DR  
XX New isolated cancer-associated polynucleotides and polypeptides useful  
PT for diagnosing, preventing or treating cancers, especially lymphoma and  
PT leukemia, or in screening for agents that modulate cancer.  
XX PT  
XX claim 16; seqid 506; 310pp; English.  
XX PS  
XX The invention relates to an isolated nucleic acid comprising at least 10  
CC contiguous nucleotides of any of the 233 polynucleotide sequences given  
CC in the specification, or its complement. The nucleic acids encode cancer-  
CC associated proteins. Also included are an expression vector comprising  
CC the isolated nucleic acid cited above, a host cell comprising the above  
CC recombinant nucleic acid or expression vector, a microarray for detecting  
CC a cancer-associated (CA) nucleic acid comprising at least one probe  
CC comprising at least 10 contiguous nucleotides of any of the above-  
CC mentioned nucleotide sequences, an isolated polypeptide (encoded within  
CC an open reading frame of a CA sequence selected from any of the 95  
CC polynucleotide sequences as mentioned in the specification, or its  
CC complement), an isolated antibody, (or its antigen binding fragment) that  
CC binds to the above polypeptide, a hybridoma that produces the above  
CC monoclonal antibody, a pharmaceutical composition comprising the above  
CC antibody and a pharmaceutical excipient, a kit for detecting cancer  
CC cells (comprising the antibody cited above, methods for diagnosing cancer  
CC or for detecting the presence or absence of cancer cells in an  
CC individual, a method for inhibiting a growth of cancer cells in an  
CC individual, a method for delivering a therapeutic agent to cancer cells  
CC in an individual, an electronic library comprising the above  
CC polynucleotide or polypeptide (or their fragments), methods of screening  
CC for anticancer activity or for a bioactive agent capable of modulating  
CC the activity of a CA protein (CAP), methods for detecting cancer  
CC associated with expression of a polypeptide in a test cell sample, a  
CC method for treating cancers and a method for inhibiting the expression of  
CC CA gene in a cell. The composition and methods are useful for detecting,  
CC diagnosing, preventing and treating cancers, especially lymphoma and  
CC leukaemia. These may also be used in screening for agents that modulate

CC cancer. The present sequence is a human CAP genomic sequence. Note: The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 326002 BP; 84319 A; 72970 C; 76829 G; 91884 T; 0 U; 0 Other;  
SQ Query Match 84.7%; Score 14.4; DB 13; Length 326002;  
Best Local Similarity 93.8%; Pred. No. 1.5e+03;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCCACGCTTGCCGAG 17  
Db 110753 CCCACGCTTGCCGAG 110768  
||||| |||||||

RESULT 82  
ABN23393/c  
ID ABN23393 standard; cDNA; 338 BP.  
XX AC ABN23393;  
XX DT 24-JUN-2002 (first entry)  
XX DE Human ORFX polynucleotide sequence SEQ ID NO:15263.  
XX KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;  
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;  
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;  
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;  
KW hypertension; hypothyroidism; cholesterol ester storage disease;  
KW immune deficiency; immune disorder; infectious disease;  
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;  
KW myasthenia gravis; gene; ss.  
XX OS Homo sapiens.  
XX PN WO200192523-A2.  
XX PD 06-DEC-2001.  
XX PF 29-MAY-2001; 2001WO-US010836.  
XX PR 30-MAY-2000; 2000US-0206132P.  
XX PR 29-AUG-2000; 2000US-0228716P.  
XX PA (CURA-) CURAGEN CORP.  
XX PI Shimkets RA, Leach MD;  
XX DR WPI; 2002-106308/14.  
XX P-PSDB; ABP07641.  
XX Novel human polypeptides and polynucleotides useful for diagnosing,  
XX preventing and treating cardiovascular disease, neurodegenerative,  
XX hyperproliferative disorders and autoimmune disorders.  
XX Disclosure; SEQ ID NO 15263; 1037pp; English.

CC The present invention describes substantially purified human proteins  
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1  
CC in the specification). ABN15762 to ABN27252 encode the human ORFX  
CC proteins given in ABP0010 to ABP1500. ORFX proteins are useful for  
CC treating or preventing a pathology associated with an ORFX-associated  
CC disorder in humans, and in the manufacture of a medicament for treating a  
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide  
CC sequences can be used in gene therapy. ORFX sequences can be used in the  
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,  
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,  
CC osteoarthritis, neurodegenerative disorders, disorders related to organ  
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic  
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester  
CC storage disease, various immune deficiencies and disorders, infectious

CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid  
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host  
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also  
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,  
CC bone degenerative disorders, or periodontal disease, and for gut  
CC protection or regeneration and treatment of lung or liver fibrosis,  
CC reperfusion injury in various tissues and conditions resulting from  
CC systemic cytokine damage. N.B. The sequence data for this patent did not  
CC form part of the printed specification, but was obtained in electronic  
CC format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 338 BP; 66 A; 98 C; 108 G; 64 T; 0 U; 2 Other;  
SQ Query Match 82.4%; Score 14; DB 6; Length 338;  
Best Local Similarity 100.0%; Pred. No. 2.2e+03;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACGCTTGCC 14  
Db 196 GCCACGCTTGCC 183  
||||| |||||||

RESULT 83  
ACH41474/c  
ID ACH41474 standard; cDNA; 458 BP.  
XX AC ACH41474;  
XX DT 13-OCT-2003 (first entry)  
XX DE Human foetal brain cDNA #2841.  
XX KW Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;  
KW genome mapping; biodiversity; genetic disorder.  
XX OS Homo sapiens.  
XX PN US2003073623-A1.  
XX PD 17-APR-2003.  
XX PF 30-JUL-2001; 2001US-00918995.  
XX PR 30-JUL-2001; 2001US-00918995.  
XX PA (DRMA/) DRMANAC R T.  
PA (LABA/) LABAT I.  
PA (STAC/) STACHE-CRAIN B.  
PA (DICK/) DICKSON M C.  
PA (JONE/) JONES L W.  
XX PI Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;  
XX WPI; 2003-615964/58.  
XX New polynucleotide sequences obtained from various cDNA libraries, useful  
XX as hybridization probes, as oligomers for PCR, for chromosome and gene  
XX mapping, in the recombinant production of protein, or in generating  
XX antisense DNA or RNA.  
XX Claim 1; SEQ ID NO 28686; 44pp; English.  
XX The invention relates to an isolated polynucleotide comprising any one of  
XX 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was  
XX determined by the technique of SBH (sequencing by hybridisation). Also  
XX included is a purified polypeptide comprising a sequence corresponding to  
XX a reading frame of the novel polynucleotide. The nucleic acid sequences  
XX are useful in diagnostics as expressed sequence tags (EST) for  
XX identifying expressed genes or for physical mapping of the human genome,  
XX in forensics, in assessing biodiversity, or in identifying mutations  
XX responsible for genetic disorders and other traits. The nucleotide  
XX sequences are also useful as hybridisation probes, as oligomers for PCR,  
XX for chromosome and gene mapping, in the recombinant production of

CC protein, or in generating antisense DNA or RNA. The purified polypeptide  
CC is useful for generating antibodies specific for it. The present sequence  
CC is one of the 38043 isolated cDNA/EST sequences. Note: The present sequence  
CC for this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html?DocID=20030073623  
XX  
SQ Sequence 458 BP; 101 A; 104 C; 124 G; 125 T; 0 U; 4 Other;  
Query Match 82.4%; Score 14; DB 9; Length 458;  
Best Local Similarity 100.0%; Pred. No. 2.2e+03;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 3 CCACGCTGGCCGA 16  
| | | | | | | | | | | | | | | |  
Db 358 CCACGCTGGCCGA 345  
RESULT 84  
ADN12918/c  
ID ADN12918 standard; cDNA; 508 BP.  
XX  
AC ADN12918;  
XX  
DT 29-JUL-2004 (first entry)  
XX  
DE Human prostate/colon/lung/breast cancer-related cDNA 433, SEQ:433.  
XX  
KW Human; cancer; tumour; prostate cancer; colon cancer; lung cancer;  
KW breast cancer; drug screening; diagnosis; prognosis; prevention;  
KW gene mapping; tissue typing; tissue profiling; cytostatic; gene therapy;  
KW ss.  
XX  
OS Homo sapiens.  
XX  
PN WO2004039943-A2.  
XX  
PD 13-MAY-2004.  
XX  
PF 16-MAY-2003; 2003WO-US015465.  
XX  
PR 17-MAY-2002; 2002US-0381533P.  
PR 04-FEB-2003; 2003US-0445222P.  
XX  
PA (CHIR ) CHIRON CORP.  
XX  
PI Scott EM, Lamson G, Kassam A, Zhang G, Sakamoto D, Garcia PD;  
XX  
DR WPI; 2004-376173/35.  
XX  
PT New isolated polynucleotides, useful for gene mapping or tissue typing or  
PT profiling, as diagnostic reagents, and for preventing or treating cancer,  
PT e.g. prostate, colon, or breast cancer.  
XX  
PS Claim 2; SEQ ID NO 433; 190pp; English.  
XX  
CC The invention relates to nucleic acids (ADN12486-ADN13970) isolated from  
CC human prostate, colon, lung and breast cancer cDNA libraries, and to 57  
CC proteins (ADN13971-ADN14027) encoded by a subset of these cDNA sequences  
CC (ADN13914-ADN13970). The invention also relates to vectors and host cells  
CC comprising a nucleic acid of the invention; a method for the recombinant  
CC production of a protein of the invention; a method for the recombinant  
CC protein of the invention; a polynucleotide library comprising at least  
CC one nucleic acid sequence of the invention; a method for detecting a  
CC cancerous cell by PCR or probe hybridisation; inhibiting a cancerous  
CC phenotype (particularly aberrant proliferation) of a cell; a method of  
CC identifying an agent that modulates the biological activity of a gene  
CC product differentially expressed in a cancerous cell compared with a  
CC normal cell; and a method of treating a cancer patient using the agent  
CC identified. The nucleic acids and polypeptides can be used to diagnose,  
CC prognose, treat or prevent cancers such as prostate, colon, lung or  
CC breast cancer, and can also be used to screen for drugs for the treatment  
CC of cancer. The nucleic acids can also be used for gene mapping, tissue

CC typing and tissue profiling. The present sequence represents a  
CC specifically claimed cancer-related cDNA of the invention. Note: The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 508 BP; 115 A; 127 C; 156 G; 110 T; 0 U; 0 Other;  
Query Match 82.4%; Score 14; DB 12; Length 508;  
Best Local Similarity 100.0%; Pred. No. 2.2e+03;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 3 CCACGCTGGCCGA 16  
| | | | | | | | | | | | | | | |  
Db 238 CCACGCTGGCCGA 225  
RESULT 85  
ADN13857/c  
ID ADN13857 standard; cDNA; 706 BP.  
XX  
AC ADN13857;  
XX  
DT 29-JUL-2004 (first entry)  
XX  
DE Human prostate/colon/lung/breast cancer-related cDNA 1372, SEQ:1372.  
XX  
KW Human; cancer; tumour; prostate cancer; colon cancer; lung cancer;  
KW breast cancer; drug screening; diagnosis; prognosis; prevention;  
KW gene mapping; tissue typing; tissue profiling; cytostatic; gene therapy;  
KW ss.  
XX  
OS Homo sapiens.  
XX  
PN WO2004039943-A2.  
XX  
PD 13-MAY-2004.  
XX  
PF 16-MAY-2003; 2003WO-US015465.  
XX  
PR 17-MAY-2002; 2002US-0381533P.  
PR 04-FEB-2003; 2003US-0445222P.  
XX  
PA (CHIR ) CHIRON CORP.  
XX  
PI Scott EM, Lamson G, Kassam A, Zhang G, Sakamoto D, Garcia PD;  
XX  
DR WPI; 2004-376173/35.  
XX  
PT New isolated polynucleotides, useful for gene mapping or tissue typing or  
PT profiling, as diagnostic reagents, and for preventing or treating cancer,  
PT e.g. prostate, colon, or breast cancer.  
XX  
PS Claim 2; SEQ ID NO 1372; 190pp; English.  
XX  
CC The invention relates to nucleic acids (ADN12486-ADN13970) isolated from  
CC human prostate, colon, lung and breast cancer cDNA libraries, and to 57  
CC proteins (ADN13971-ADN14027) encoded by a subset of these cDNA sequences  
CC (ADN13914-ADN13970). The invention also relates to vectors and host cells  
CC comprising a nucleic acid of the invention; a method for the recombinant  
CC production of a protein of the invention; a method for the recombinant  
CC protein of the invention; a polynucleotide library comprising at least  
CC one nucleic acid sequence of the invention; a method for detecting a  
CC cancerous cell by PCR or probe hybridisation; inhibiting a cancerous  
CC phenotype (particularly aberrant proliferation) of a cell; a method of  
CC identifying an agent that modulates the biological activity of a gene  
CC product differentially expressed in a cancerous cell compared with a  
CC normal cell; and a method of treating a cancer patient using the agent  
CC identified. The nucleic acids and polypeptides can be used to diagnose,  
CC prognose, treat or prevent cancers such as prostate, colon, lung or  
CC breast cancer, and can also be used to screen for drugs for the treatment  
CC of cancer. The nucleic acids can also be used for gene mapping, tissue  
CC typing and tissue profiling. The present sequence represents a



CC specifically claimed cancer-related cDNA of the invention. Note: The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 706 BP; 170 A; 176 C; 206 G; 150 T; 0 U; 4 Other;  
Query Match 82.4%; Score 14; DB 12; Length 706;  
Best Local Similarity 100.0%; Pred. No. 2.2e+03; Mismatches 0; Indels 0; Gaps 0;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 3 CCACGCTTGGCCGA 16  
Db 238 CCACGCTTGGCCGA 225  
|||||

## RESULT 86

AAC47788

ID AAC47788 standard; DNA; 1234 BP.

XX AAC47788;

AC AAC47788;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 55122.

KW Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway; metabolic pathway;  
KW promoter; termination sequence; ss.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

XX 25-FEB-2000; 2000EP-00301439.

PR 25-FEB-1999; 99US-0121825P.

PR 05-MAR-1999; 99US-0123180P.

PR 09-MAR-1999; 99US-0123548P.

PR 23-MAR-1999; 99US-0125788P.

PR 25-MAR-1999; 99US-0126264P.

PR 29-MAR-1999; 99US-0126785P.

PR 01-APR-1999; 99US-0127462P.

PR 06-APR-1999; 99US-0128234P.

PR 08-APR-1999; 99US-0128714P.

PR 16-APR-1999; 99US-0129845P.

PR 19-APR-1999; 99US-0130077P.

PR 21-APR-1999; 99US-0130449P.

PR 23-APR-1999; 99US-0130891P.

PR 28-APR-1999; 99US-0131449P.

PR 30-APR-1999; 99US-0132048P.

PR 30-APR-1999; 99US-0132407P.

PR 04-MAY-1999; 99US-0132484P.

PR 05-MAY-1999; 99US-0132485P.

PR 06-MAY-1999; 99US-0132486P.

PR 06-MAY-1999; 99US-0134487P.

PR 07-MAY-1999; 99US-0132863P.

PR 11-MAY-1999; 99US-0134256P.

PR 14-MAY-1999; 99US-0134218P.

PR 14-MAY-1999; 99US-0134219P.

PR 14-MAY-1999; 99US-0134221P.

PR 14-MAY-1999; 99US-0134370P.

PR 18-MAY-1999; 99US-0134768P.

PR 19-MAY-1999; 99US-0134941P.

PR 20-MAY-1999; 99US-0135124P.

PR 21-MAY-1999; 99US-0135353P.

PR 24-MAY-1999; 99US-0135629P.

PR 25-MAY-1999; 99US-0136021P.

PR 27-MAY-1999; 99US-0136392P.

PR 28-MAY-1999; 99US-0136782P.

PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140351P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142053P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.

PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149388P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 20-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 23-AUG-1999; 99US-0150566P.  
PR 25-AUG-1999; 99US-0150884P.  
PR 26-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158039P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161992P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 82.4%; Score 14; DB 3; Length 1234;  
Best Local Similarity 100.0%; Pred. No. 2.3e+03;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCC 14  
Db 563 GCCCAGCTTGGCC 576

RESULT 87

ADN46845\_16/c  
Continuation (17 of 21) of ADN46845 from base 1600001 (Thermococcus kodakaraensis KOD1 Gen  
WP Sequence split into 21 fragments LOCUS ADN46845 Accession Adn46845

WP	Fragment Name	Begin	End
WP	ADN46845_00	1	110000
WP	ADN46845_01	100001	210000
WP	ADN46845_02	200001	310000
WP	ADN46845_03	300001	410000
WP	ADN46845_04	400001	510000
WP	ADN46845_05	500001	610000
WP	ADN46845_06	600001	710000
WP	ADN46845_07	700001	810000
WP	ADN46845_08	800001	910000
WP	ADN46845_09	900001	1010000
WP	ADN46845_10	1000001	1110000
WP	ADN46845_11	1100001	1210000
WP	ADN46845_12	1200001	1310000
WP	ADN46845_13	1300001	1410000
WP	ADN46845_14	1400001	1510000
WP	ADN46845_15	1500001	1610000
WP	ADN46845_16	1600001	1710000
WP	ADN46845_17	1700001	1810000
WP	ADN46845_18	1800001	1910000
WP	ADN46845_19	1900001	2010000
WP	ADN46845_20	2000001	2089378

Query Match 82.4%; Score 14; DB 12; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 2.4e+03;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 CCACGCTTGGCCGA 16  
Db 53895 CCACGCTTGGCCGA 53882

RESULT 88

ADN47591\_04  
Continuation (5 of 21) of ADN47591 from base 400001 (Thermococcus kodakaraensis KOD1 Gen  
WP Sequence split into 21 fragments LOCUS ADN47591 Accession Adn47591

WP	Fragment Name	Begin	End
WP	ADN47591_00	1	110000
WP	ADN47591_01	100001	210000
WP	ADN47591_02	200001	310000
WP	ADN47591_03	300001	410000
WP	ADN47591_04	400001	510000
WP	ADN47591_05	500001	610000
WP	ADN47591_06	600001	710000
WP	ADN47591_07	700001	810000
WP	ADN47591_08	800001	910000
WP	ADN47591_09	900001	1010000
WP	ADN47591_10	1000001	1110000
WP	ADN47591_11	1100001	1210000
WP	ADN47591_12	1200001	1310000
WP	ADN47591_13	1300001	1410000
WP	ADN47591_14	1400001	1510000
WP	ADN47591_15	1500001	1610000
WP	ADN47591_16	1600001	1710000
WP	ADN47591_17	1700001	1810000
WP	ADN47591_18	1800001	1910000
WP	ADN47591_19	1900001	2010000
WP	ADN47591_20	2000001	2089378

Query Match 82.4%; Score 14; DB 12; Length 110000;

Best Local Similarity 100.0%; Pred. No. 2.4e+03;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 CCACGCTTGGCCGA 16  
|||||  
DB 35483 CCACGCTTGGCCGA 35496

## RESULT 89

ADN46123\_16/c  
Continuation (17 of 21) of ADN46123 from base 1600001 (Thermococcus kodakaraensis KOD1 gen  
WP Sequence split into 21 fragments LOCUS ADN46123 Accession Adn46123

Fragment Name	Begin	End
WP ADN46123_00	1	110000
WP ADN46123_01	100001	210000
WP ADN46123_02	200001	310000
WP ADN46123_03	300001	410000
WP ADN46123_04	400001	510000
WP ADN46123_05	500001	610000
WP ADN46123_06	600001	710000
WP ADN46123_07	700001	810000
WP ADN46123_08	800001	910000
WP ADN46123_09	900001	1010000
WP ADN46123_10	1000001	1110000
WP ADN46123_11	1100001	1210000
WP ADN46123_12	1200001	1310000
WP ADN46123_13	1300001	1410000
WP ADN46123_14	1400001	1510000
WP ADN46123_15	1500001	1610000
WP ADN46123_16	1600001	1710000
WP ADN46123_17	1700001	1810000
WP ADN46123_18	1800001	1910000
WP ADN46123_19	1900001	2010000
WP ADN46123_20	2000001	2089378

Query Match 82.4%; Score 14; DB 12; Length 110000;

Best Local Similarity 100.0%; Pred. No. 2.4e+03;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 CCACGCTTGGCCGA 16  
|||||  
DB 53895 CCACGCTTGGCCGA 53882

## RESULT 90

ADN47209\_04  
Continuation (5 of 21) of ADN47209 from base 400001 (Thermococcus kodakaraensis KOD1 gen  
WP Sequence split into 21 fragments LOCUS ADN47209 Accession Adn47209

Fragment Name	Begin	End
WP ADN47209_00	1	110000
WP ADN47209_01	100001	210000
WP ADN47209_02	200001	310000
WP ADN47209_03	300001	410000
WP ADN47209_04	400001	510000
WP ADN47209_05	500001	610000
WP ADN47209_06	600001	710000
WP ADN47209_07	700001	810000
WP ADN47209_08	800001	910000
WP ADN47209_09	900001	1010000
WP ADN47209_10	1000001	1110000
WP ADN47209_11	1100001	1210000
WP ADN47209_12	1200001	1310000
WP ADN47209_13	1300001	1410000
WP ADN47209_14	1400001	1510000
WP ADN47209_15	1500001	1610000
WP ADN47209_16	1600001	1710000
WP ADN47209_17	1700001	1810000
WP ADN47209_18	1800001	1910000
WP ADN47209_19	1900001	2010000
WP ADN47209_20	2000001	2089378

Query Match

Best Local Similarity 100.0%; Pred. No. 2.4e+03;  
Matches 14; DB 12; Length 110000;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 CCACGCTTGGCCGA 16  
|||||  
DB 35483 CCACGCTTGGCCGA 35496

## RESULT 91

ADN46464\_16/c  
Continuation (17 of 21) of ADN46464 from base 1600001 (Thermococcus kodakaraensis KOD1 gen  
WP Sequence split into 21 fragments LOCUS ADN46464 Accession Adn46464

Fragment Name	Begin	End
WP ADN46464_00	1	110000
WP ADN46464_01	100001	210000
WP ADN46464_02	200001	310000
WP ADN46464_03	300001	410000
WP ADN46464_04	400001	510000
WP ADN46464_05	500001	610000
WP ADN46464_06	600001	710000
WP ADN46464_07	700001	810000
WP ADN46464_08	800001	910000
WP ADN46464_09	900001	1010000
WP ADN46464_10	1000001	1110000
WP ADN46464_11	1100001	1210000
WP ADN46464_12	1200001	1310000
WP ADN46464_13	1300001	1410000
WP ADN46464_14	1400001	1510000
WP ADN46464_15	1500001	1610000
WP ADN46464_16	1600001	1710000
WP ADN46464_17	1700001	1810000
WP ADN46464_18	1800001	1910000
WP ADN46464_19	1900001	2010000
WP ADN46464_20	2000001	2089378

Query Match 82.4%; Score 14; DB 12; Length 110000;

Best Local Similarity 100.0%; Pred. No. 2.4e+03;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 CCACGCTTGGCCGA 16  
|||||  
DB 53895 CCACGCTTGGCCGA 53882

## RESULT 92

ADN47960\_04  
Continuation (5 of 21) of ADN47960 from base 400001 (Thermococcus kodakaraensis KOD1 gen  
WP Sequence split into 21 fragments LOCUS ADN47960 Accession Adn47960

Fragment Name	Begin	End
WP ADN47960_00	1	110000
WP ADN47960_01	100001	210000
WP ADN47960_02	200001	310000
WP ADN47960_03	300001	410000
WP ADN47960_04	400001	510000
WP ADN47960_05	500001	610000
WP ADN47960_06	600001	710000
WP ADN47960_07	700001	810000
WP ADN47960_08	800001	910000
WP ADN47960_09	900001	1010000
WP ADN47960_10	1000001	1110000
WP ADN47960_11	1100001	1210000
WP ADN47960_12	1200001	1310000
WP ADN47960_13	1300001	1410000
WP ADN47960_14	1400001	1510000
WP ADN47960_15	1500001	1610000
WP ADN47960_16	1600001	1710000
WP ADN47960_17	1700001	1810000
WP ADN47960_18	1800001	1910000
WP ADN47960_19	1900001	2010000
WP ADN47960_20	2000001	2089378

Query Match

Best Local Similarity 100.0%; Pred. No. 2.4e+03;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	3	CCACGCTTGCCCGA 16	Matches	14;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps
Db	35483	CCACGCTTGCCCGA 35496									
RESULT 93											
ABQ81848/c											
ID	ABQ81848	standard; DNA; 349980 BP.									
XX	AC	ABQ81848;									
XX	DT	19-NOV-2002 (first entry)									
XX	XX	Bifidobacterium longum NCC2705 related nucleotide sequence SEQ ID:1104.									
XX	XX	Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;									
KW	KW	antidiarrhetic; antibacterial; inhibitor of salmonella; detection;									
KW	KW	identification; lactic acid bacterium; diarrhoea; pathogenic bacteria;									
KW	KW	rotavirus; food composition; pharmaceutical composition; gene; ds.									
XX	XX	Bifidobacterium longum.									
OS	OS	Synthetic.									
XX	XX	EP1227152-A1.									
XX	XX	31-JUL-2002.									
XX	XX	30-JAN-2001; 2001EP-00102050.									
XX	XX	30-JAN-2001; 2001EP-00102050.									
XX	XX	(NEST ) SOC PROD NESTLE SA.									
XX	XX	WPI; 2002-668397/72.									
XX	XX	Novel polynucleotide comprising Bifidobacterium genome sequence useful as a probe or primer for detecting and/or identifying Bifidobacterium longum in a biological sample.									
PT	PT	Disclosure; SEQ ID NO 1104; 80pp; English.									
PS	PS	The present invention describes a polynucleotide (I) comprising a sequence of a Bifidobacterium genome selected from the nucleotide sequences given in ABQ81842 and ABQ81843, or a sequence exhibiting at least 90% identity or which hybridises with the sequences given in ABQ81842 and ABQ81843. Also described is a polynucleotide (II) encoding a fusion protein, comprising a sequence selected from 1097 sequences given in ABP65258 to ABP66354 ligated in frame to a polynucleotide encoding a heterologous polypeptide. (I) has antidiarrhetic and antibacterial activities, and can be used as an inhibitor of Salmonella. (I) (which is a probe) is useful for the detection and/or identification of Bifidobacterium longum in a biological sample. A carrier containing the lactic acid bacterium Bifidobacterium longum NCC2705 (CNCM I-2618) can be used for preventing and/or treating diarrhoea brought about by pathogenic bacteria and/or rotavirus. The carrier is a food composition selected from milk, yogurt, curd, cheese, fermented milks, milk based fermented products, ice-creams, fermented cereal based products, milk based powders, infant formula, pet food or a pharmaceutical composition selected from tablets, liquid bacterial suspensions, dried oral supplement, wet oral supplement, dry tube feeding or wet tube feeding. (I) is useful in DNA arrays or chips to carry out analysis of the expression of the Bifidobacterium gene. ABQ81844 to ABQ81850 represent Bifidobacterium related nucleotide sequences given in the Sequence Listing from the present invention but not mentioned further within the specification. N.B. The sequence data for this patent is not represented in the printed specification but is based on sequence information supplied by the European Patent Office									
XX	XX	Sequence 349980 BP; 69195 A; 106952 C; 106128 G; 67705 T; 0 U; 0 Other;									
XX	XX	Query Match									
XX	XX	Best Local Similarity 100.0%; Pred. No. 2.4e-03;									
XX	XX	82.4%; Score 14; DB 6; Length 349980;									
XX	XX	Query Match									
XX	XX	Best Local Similarity 100.0%; Pred. No. 2.4e-03;									
XX	XX	82.4%; Score 14; DB 6; Length 349980;									
XX	XX	Query Match									
XX	XX	Best Local Similarity 100.0%; Pred. No. 2.4e-03;									
XX	XX	82.4%; Score 14; DB 6; Length 349980;									
XX	XX	Query Match									
XX	XX	Best Local Similarity 100.0%; Pred. No. 2.4e-03;									
XX	XX	82.4%; Score 14; DB 6; Length 349980;									
XX	XX	Query Match									
XX	XX	Best Local Similarity 100.0%; Pred. No. 2.4e-03;									
XX	XX	82.4%; Score 14; DB 6; Length 349980;									
XX	XX	Query Match									
XX	XX	Best Local Similarity 100.0%; Pred. No. 2.4e-03;									
XX	XX	82.4%; Score 14; DB 6; Length 349980;	</								

```

Best Local Similarity 100.0%; Pred. No. 2.4e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCCAGGCTTGCC 14
Db 242514 GCCCAGGCTTGCC 242527

RESULT 95
AAZ68079
ID AAZ68079 standard; DNA; 47 BP.
XX
AC AAZ68079;
XX
XX 10-SEP-2001 (first entry)
XX
DE Human map-related biallelic marker SEQ ID NO:2426.
XX
XX Human genome; biallelic marker; high density disequilibrium map;
KW genomic map; haplotype; phenotype; polymorphic base; genotyping;
KW haplotyping; hybridisation; identification; characterisation; diagnosis;
KW single nucleotide polymorphism; SNP; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT variation replace(24,C)
FT /*tag= a
FT /standard_name= "single nucleotide polymorphism"
XX
XX WO9954500-A2.
XX
XX 28-OCT-1999.
XX
XX 21-APR-1999; 99WO-IB000822.
XX
XX 21-APR-1998; 98US-0082614P.
XX
XX 23-NOV-1998; 98US-0109732P.
XX
XX (GEST ) GENSET.
XX
XX Cohen D, Blumenfeld M, Chumakov I;
XX
XX WPI; 2000-013267/01.
XX
XX Novel biallelic markers used to construct a high density disequilibrium
XX map of the human genome.
XX
XX Claim 3; Page 748; 2745pp; English.
XX
XX AAZ65654 to AAZ69578 represent human biallelic markers from the present
XX invention, which contain a polymorphic base at position 24 of their
XX nucleotide sequences. AAZ69579 to AAZ77440 represent amplification
XX primers for the biallelic markers. The biallelic markers of the invention
XX have a variety of uses: they can be used for high density mapping of the
XX human genome, and in complex association studies and haplotyping studies
XX which are useful in determining the genetic basis for disease states.
XX Compositions and methods of the invention can also be useful for the
XX identification of the targets for the development of pharmaceutical
XX agents and diagnostic methods, as well as the characterisation of the
XX differential efficacious responses to and side effects from
XX pharmaceutical agents acting on a disease as well as other treatment.
XX N.B. The SEQ ID NOS 2852, 2913, 2974, 3035, 3096, 3157, 3227, 3297 and
XX 3367, are not actually given a sequence in the Sequence Listing from the
XX present invention
XX
XX Sequence 47 BP; 8 A; 17 C; 11 G; 11 T; 0 U; 0 Other;

Query Match 81.2%; Score 13.8; DB 3; Length 47;
Best Local Similarity 88.2%; Pred. No. 2.7e+03;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCCCAGGCTTGCCGAG 17
Db 27 GCCCAGGCTTGTCAGAG 43

RESULT 96
ABN78126
ID ABN78126 standard; cDNA; 228 BP.
XX
XX ABN78126;
AC
XX
XX 08-JUL-2002 (first entry)
DT
XX
DE Human glycoprotein-like ORF3073 cDNA, SEQ ID NO:6145.
XX
XX Human; ORF; open reading frame; ORFX; drug screening; diagnosis;
KW disease monitoring; cytokine; cell proliferation; cell differentiation;
KW immune modulation; haematopoiesis regulation; tissue growth;
KW angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; tumour inhibition; bodily characteristic; fertility;
KW behaviour; cancer; proliferative disorder; neurological disorder;
KW cardiovascular disease; immune system disorder; organ transplantation;
KW tissue growth disorder; tissue regeneration disorder; diabetes mellitus;
KW hypothyroidism; cholesterol ester storage disease; infection; vulnery;
KW vasotropic; antipsoriatic; antidiabetic; cyostatic; neurotropic;
KW neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;
KW cardiant; hypotensive; antithyroid; antiinflammatory; immunomodulator;
KW dermatological; analgesic; virucide; antibacterial; fungicide; gene; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200190366-A2.
XX
XX 29-NOV-2001.
PD
XX
XX 24-MAY-2001; 2001WO-US017076.
PF
XX
XX 24-MAY-2000; 2000US-0206690P.
PR
XX
XX (CURA-) CURAGEN CORP.
PA
XX
XX Leach MD, Shimkets RA;
PI
XX
XX WPI; 2002-106200/14.
DR
XX
XX P-PSDB; ABP34100.
DR
XX
XX Novel human polypeptides and polynucleotides useful for diagnosing,
XX preventing and treating cardiovascular disease, neurodegenerative,
XX hyperproliferative disorders and disorders related to organ
XX transplantation.
XX
XX Claim 1; Page 1786; 2508pp; English.
XX
XX Sequences ABP31028-ABP35561 represent 4334 novel human proteins
XX designated ORF (open reading frame) 1-4534, and sequences ABN75054-
XX ABN79587 represent cDNAs encoding them. The invention also encompasses
XX polypeptides at least 80% identical to the ORF1-ORF4534 (collectively
XX referred to as ORFX) proteins, polynucleotides at least 85% identical to
XX the ORFX nucleic acid sequences, vectors and host cells comprising ORFX
XX polynucleotides, the recombinant production of ORFX proteins, antibodies
XX specific for ORFX proteins, methods of detecting ORFX polynucleotides and
XX polypeptides, methods of screening for modulators of ORFX expression or
XX activity, and methods of screening individuals for a predisposition to an
XX ORFX-associated disorder. The ORFX proteins of the invention have a wide
XX range of biological activities, such as cytokine, cell proliferation,
XX cell differentiation, immune modulation, haematopoiesis regulation,
XX tissue growth, angiogenesis, activin or inhibin activity, chemotactic/
XX chemokinetic activity, haemostatic activity, tumour inhibition activity,
XX receptor/ligand, antiinflammatory activity, tumour inhibition activity,
XX and antiinfective activity, and may also be involved in the determination
XX of bodily characteristics, fertility and behaviour. ORFX proteins,
XX nucleic acids and antibodies may be used in the treatment of cancers,
XX other proliferative disorders such as psoriasis and benign tumours,
XX neurological disorders such as epilepsy and Alzheimer's disease,

```

CC cardiovascular diseases, immune system disorders, disorders related to  
CC organ transplantation, disorders of tissue growth and regeneration,  
CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol  
CC storage disease, and infectious diseases caused by viral, bacterial,  
CC fungal and other pathogens. ORFX nucleic acids may also be used as a  
CC source of primers and probes in the detection of ORFX genomic sequences  
CC or transcripts, in the identification and cloning of homologous  
CC sequences, in genetic diagnosis, and in forensic biology. The ORFX  
CC nucleic acids may additionally be used to produce transgenic animals  
CC which may be useful for studying the function and/or activity of ORFX  
CC protein, and in drug screening. The ORFX proteins may also be used as  
CC immunogens to generate specific antibodies, which are useful in the  
CC diagnosis, treatment and monitoring of ORFX-associated diseases  
XX  
SQ Sequence 228 BP; 42 A; 77 C; 80 G; 28 T; 0 U; 1 Other;  
Query Match 81.2%; Score 13.8; DB 6; Length 228;  
Best Local Similarity 88.2%; Pred. No. 2.8e+03;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 GCCCAGCGCTTGGCCGAG 17  
Db 65 GCGCAGCGCTTGGCCGAG 81  
RESULT 97  
AAD48755 ID AAD48755 standard; RNA; 247 BP.  
XX AAD48755;  
XX 07-MAR-2003 (first entry)  
XX Human acyl CoA:diacylglycerol acetyltransferase (DGAT) RNA fragment #1.  
XX Human; RNA binding molecule; acyl CoA:diacylglycerol acetyltransferase;  
XX DGAT; ss.  
XX Homo sapiens.  
XX WO200281748-A1.  
XX 17-OCT-2002.  
XX 04-APR-2002; 2002WO-SE000677.  
XX 05-APR-2001; 2001SE-00001218.  
XX 05-APR-2001; 2001US-0281384P.  
XX (BIOV-) BIOVITRUM AB.  
XX Ekblom J;  
XX WPI; 2003-058568/05.  
XX Identifying RNA-binding molecule by predicting structure of RNA fragment,  
XX synthesizing DNA fragment corresponding to predicted RNA structure,  
XX performing reporter gene assay after placing the DNA upstream of reporter  
XX gene.  
XX Claim 12; Page 27; 35pp; English.  
XX The present invention relates to a method of identifying RNA-binding  
CC molecule comprising predicting the structure of RNA-fragment, selecting  
CC suitable predicted RNA fragment with an individual stem, synthesising a  
CC DNA-fragment corresponding to the RNA fragment, inserting the DNA  
CC fragment in upstream proximity of reporter assay gene to form reporter  
CC construct and performing a reporter gene assay which detects interaction  
CC between a molecule to be tested for RNA-binding and RNA fragment of the  
CC reporter construct. The method is useful for identifying an RNA binding  
CC molecule. The present sequence is human acyl CoA:diacylglycerol-  
CC acetyltransferase (DGAT) RNA fragment. This sequence is used to  
CC illustrate the method of the invention

XX SQ Sequence 247 BP; 32 A; 74 C; 108 G; 0 T; 33 U; 0 Other;  
Query Match 81.2%; Score 13.8; DB 10; Length 247;  
Best Local Similarity 76.5%; Pred. No. 2.8e+03;  
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 GCCCAGCGCTTGGCCGAG 17  
Db 208 GCCCAGCGCTTGGCCGAG 224  
RESULT 98  
ABX84111/c  
ID ABX84111 standard; cDNA; 283 BP.  
XX ABX84111;  
XX 24-APR-2003 (first entry)  
XX Corn ear-derived polynucleotide (cpd) #2571.  
XX Corn ear-derived polynucleotide; cdp; cDNA library; SATMON022; SATMON023;  
XX structural gene; functional gene; regulatory gene;  
XX corn ear-specific profile; gene transcription; gene expression;  
XX hybrid plant; desirable trait expression; plant breeding program;  
XX inheritance; desired characteristic; growth; development;  
XX disease resistance; environmental adaptability; quality; yield;  
XX multigene trait; plant; gene; ss.  
XX Zea mays.  
XX US6476212-B1.  
XX 05-NOV-2002.  
XX 14-MAY-1999; 99US-00313294.  
XX 26-MAY-1998; 98US-0086722P.  
XX (INCY-) INCYTE GENOMICS INC.  
XX Lalgudi RV, Ito LY, Sherman BK;  
XX WPI; 2003-208840/20.  
XX Novel purified corn-ear derived polynucleotide useful as hybridization  
XX probe for detecting polynucleotide in sample, and for identifying,  
XX evaluating, and altering desired characteristics associated with growth,  
XX development.  
XX Example; SEQ ID NO 2571; 390pp; English.  
XX The present invention relates to the isolation of corn ear-derived  
XX polynucleotides (cdps) from the corn (Zea mays) cDNA libraries SATMON022  
XX and SATMON023. Some of the cdps uniquely identify structural, functional,  
XX and regulatory genes of corn ear. The polynucleotides sequences are  
XX useful for detecting cdps in a sample, for producing a corn ear-specific  
XX profile of gene transcription, for detecting altered gene expression in  
XX inbred or hybrid plants, and for screening several molecules for specific  
XX binding to the polynucleotide. The cdps are useful to identify, isolate,  
XX or extend identical or related corn-ear nucleic acid sequences from DNA  
XX libraries, and in nucleic acid amplification or hybridisation techniques  
XX to follow the expression of desirable traits through plant breeding  
XX programs. Preferably, the cdps are used to identify, evaluate, alter, or  
XX follow the inheritance of desired characteristics associated with growth  
XX and development, disease resistance, environmental adaptability, quality,  
XX and yield of corn. The cdps are also useful as molecular markers for  
XX studying inheritance and multigene traits in a plant breeding program.  
XX The cdps are useful for producing purified corn-ear polypeptides by  
XX recombinant techniques. They are also useful in diagnostic assays to  
XX detect or confirm conditions or diseases associated with abnormal levels  
XX of cdp expression. ABX81541-ABX89140 represent corn ear-derived

CC polynucleotides (cpds) of the invention. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from the USPTO web site at  
CC seqdata.uspto.gov/psipidsEntry.html

XX Sequence 283 BP; 82 A; 56 C; 73 G; 72 T; 0 U; 0 Other;  
SQ  
Query Match 81.2%; Score 13.8; DB 10; Length 283;  
Best Local Similarity 88.2%; Pred. No. 2.8e+03;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 GCCCAGCTTGCCGAG 17  
| | | | | | | | | | | | | | | | |  
Db 106 GCCCACCCTTGCCCAAG 90

RESULT 99  
ADP95237/c  
ID ADP95237 standard; cDNA; 290 BP.  
XX  
AC ADP95237;  
XX  
DT 09-SEP-2004 (first entry)  
XX  
DE Cotton expressed sequence tag, EST, #4248.  
XX  
KW Cotton; ss; EST; expressed sequence tag; plant; plant protection;  
KW plant improvement; marker-assisted breeding.  
XX  
OS Gossypium hirsutum; variety Nutcoton33B.  
XX  
PN US2004123338-A1.  
XX  
PD 24-JUN-2004.  
XX  
PF 08-DEC-2000; 2000US-00732627.  
XX  
PR 10-DEC-1999; 99US-0170255P.  
XX  
PA (FINC/) FINCHER K L.  
XX  
PI Fincher KL;  
XX  
PS WPI; 2004-479807/45.  
XX  
PT New substantially purified nucleic acid molecule that encodes a cotton  
PT protein or its fragment, useful as molecular tool for the targeting and  
PT isolation of novel genes for plant protection and improvement.  
XX  
PS Claim 1; SEQ ID NO 4248; 30pp; English.

XX The invention relates to a substantially purified nucleic acid molecule  
CC that encodes a cotton protein or its fragment comprising an EST  
CC (expressed sequence tag) appearing as ADP9990-ADP95919. Also included  
CC are a substantially purified cotton protein or its fragment encoded by a  
CC nucleic acid molecule above and a transformed plant (having a nucleic  
CC acid molecule which comprises: an exogenous promoter region which  
CC functions in a plant cell to cause the production of a mRNA molecule; a  
CC structural nucleic acid molecule comprising one of the ESTs or their  
CC complements; a 3' non-translated sequence that functions in the plant  
CC cell to cause termination of transcription and addition of polyadenylated  
CC ribonucleotides to a 3' end of the mRNA molecule). The ESTs are useful as  
CC molecular tool for the targeting and isolation of novel genes for plant  
CC protection and improvement. The ESTs are useful for developing new  
CC strategies for understanding critical plant developmental and metabolic  
CC pathways, for isolating genes and promoters, for identifying and mapping  
CC the genes involved in developmental and metabolic pathways, and for  
CC determining gene function. The cotton nucleic acid molecules are useful  
CC as molecular tags to isolate genetic regions, isolate genes, map genes,  
CC and determine gene function. The nucleic acid molecules are useful for  
CC determining if genes are members of a particular gene family and for use  
CC in marker-assisted breeding programs. The present sequence is one of the  
CC 4930 cotton ESTs of the invention. NOTE: The sequences are not displayed

CC in the specification but are available in electronic format from the  
CC USPTO at seqdata.uspto.gov/sequence.html?DocID=20040123338.

XX Sequence 290 BP; 81 A; 63 C; 67 G; 79 T; 0 U; 0 Other;  
SQ  
Query Match 81.2%; Score 13.8; DB 12; Length 290;  
Best Local Similarity 88.2%; Pred. No. 2.8e+03;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 GCCCAGCTTGCCGAG 17  
| | | | | | | | | | | | | | | | |  
Db 116 GCCCAGCTTGCCGAG 100

RESULT 100  
ABX86649/c  
ID ABX86649 standard; cDNA; 293 BP.  
XX  
AC ABX86649;  
XX  
DT 24-APR-2003 (first entry)  
XX  
DE Corn ear-derived polynucleotide (cpd) #5109.  
XX  
KW Corn ear-derived polynucleotide; cdp; cDNA library; SATMON022; SATMON023;  
KW structural gene; functional gene; regulatory gene;  
KW corn ear-specific profile; gene transcription; gene expression;  
KW hybrid plant; desirable trait expression; plant breeding program;  
KW inheritance; desired characteristic; growth; development;  
KW disease resistance; environmental adaptability; quality; yield;  
KW multigene trait; plant; gene; ss.  
XX  
OS Zea mays.  
XX  
PN US6476212-B1.  
XX  
PD 05-NOV-2002.  
XX  
PF 14-MAY-1999; 99US-00313294.  
XX  
PR 26-MAY-1998; 99US-0086722P.  
XX  
PA (INCY-) INCYTE GENOMICS INC.  
XX  
PI Lalgudi RV, Ito LY, Sherman BK;  
XX  
PS WPI; 2003-208840/20.  
XX  
PT Novel purified corn-ear derived polynucleotide useful as hybridization  
PT probe for detecting polynucleotide in sample, and for identifying,  
PT evaluating, and altering desired characteristics associated with growth,  
PT development.  
XX  
PS Example; SEQ ID NO 5109; 390pp; English.

XX The present invention relates to the isolation of corn ear-derived  
CC polynucleotides (cdps) from the corn (Zea mays) cDNA libraries SATMON022  
CC and SATMON023. Some of the cdps uniquely identify structural, functional,  
CC and regulatory genes of corn ear. The polynucleotide sequences are  
CC useful for detecting cdps in a sample, for producing a corn ear-specific  
CC profile of gene transcription, for detecting altered gene expression in  
CC inbred or hybrid plants, and for screening several molecules for specific  
CC binding to the polynucleotide. The cdps are useful to identify, isolate,  
CC or extend identical or related corn-ear nucleic acid sequences from DNA  
CC libraries, and in nucleic acid amplification or hybridisation techniques  
CC to follow the expression of desirable traits through plant breeding  
CC programs. Preferably, the cdps are used to identify, evaluate, alter, or  
CC follow the inheritance of desired characteristics associated with growth  
CC and development, disease resistance, environmental adaptability, quality,  
CC and yield of corn. The cdps are also useful as molecular markers for  
CC studying inheritance and multigene traits in a plant breeding program.  
CC The cdps are useful for producing purified corn-ear polypeptides by  
CC recombinant techniques. They are also useful in diagnostic assays to

CC detect or confirm conditions or diseases associated with abnormal levels  
 CC of cdp expression. ABX81541-ABX89140 represent corn ear-derived  
 CC polynucleotides (cpds) of the invention. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from the USPTO web site at  
 CC seqdata.uspto.gov/paipsDIDEntry.html

XX  
 SQ Sequence 293 BP; 77 A; 62 C; 72 G; 75 T; 0 U; 7 Other;

Query Match 81.2%; Score 13.8; DB 10; Length 293;  
 Best Local Similarity 88.2%; Pred. No. 2.8e+03;  
 Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GCCCAGCGCTTGGCCGAG 17  
 ||||| ||||| ||  
 Db 180 GCCCAGCGCTTGGCCCAAG 164

Search completed: June 10, 2006, 15:34:01  
 Job time : 160.617 secs



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2006, 15:19:28 ; Search time 2097.8 Seconds  
(without alignments)  
453.155 Million cell updates/sec

Title: US-10-600-816-31  
Perfect score: 17  
Sequence: 1 gccacgctggccgag 17

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 48236798 seqs, 2795965780 residues

Total number of hits satisfying chosen parameters: 96473596

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 1000 summaries

Database :

EST: \*  
1: gb\_est1: \*  
2: gb\_est3: \*  
3: gb\_est4: \*  
4: gb\_est5: \*  
5: gb\_est6: \*  
6: gb\_est7: \*  
7: gb\_est8: \*  
8: gb\_est9: \*  
9: gb\_est10: \*  
10: gb\_est11: \*  
11: gb\_est12: \*  
12: gb\_est13: \*  
13: gb\_est14: \*  
14: gb\_est15: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	17	100.0	158	2	B1063311	IL3-UT011
C 2	17	100.0	307	7	BE818833	CM2-BN030
C 3	17	100.0	341	7	BE818817	CM2-BN030
C 4	17	100.0	367	1	AA112374	zn68e10.r
C 5	17	100.0	384	7	BE818835	CM2-BN030
C 6	17	100.0	396	7	BE818828	CM2-BN030
C 7	17	100.0	404	7	BE818840	CM2-BN030
C 8	17	100.0	424	7	BE818805	CM2-BN030
C 9	17	100.0	453	5	CD672930	fg18a05.Y
C 10	17	100.0	458	7	BE818871	CM2-BN030
C 11	17	100.0	466	7	BE818859	CM2-BN030
C 12	17	100.0	467	7	BE818852	CM2-BN030
C 13	17	100.0	528	7	AW390096	CM3-ST017
C 14	17	100.0	571	3	BP253378	BP253378
C 15	17	100.0	577	4	CB152636	K-EST0209
C 16	17	100.0	578	3	BP267556	BP267556
C 17	17	100.0	580	3	BP265234	BP265234
C 18	17	100.0	580	3	BP267627	BP267627
C 19	17	100.0	580	3	BP268151	BP268151

20	17	100.0	582	3	BP256144	
21	17	100.0	582	3	BP267918	
22	17	100.0	582	3	BP363758	
23	17	100.0	583	7	BE006062	
24	17	100.0	584	3	BP256147	
25	17	100.0	621	4	CB130950	
26	17	100.0	630	7	BE747107	
27	17	100.0	637	3	BM843051	
28	17	100.0	699	7	BF568108	
29	17	100.0	702	2	BG251131	
30	17	100.0	851	3	BQ887352	
31	17	100.0	865	4	CA454746	
32	17	100.0	877	2	BM011054	
33	17	100.0	899	3	BQ598341	
34	17	100.0	901	2	BG831564	
35	17	100.0	916	5	CF780868	
C	36	17	100.0	918	4	BX372160
37	17	100.0	923	3	BX543952	
38	17	100.0	962	4	BX370558	
39	17	100.0	1023	3	BQ067433	
40	17	100.0	1125	2	BI193620	
41	17	100.0	1830	6	CR597125	
C	42	15.4	322	1	AA041289	
C	43	15.4	363	4	BY153767	
C	44	15.4	365	4	BY774337	
C	45	15.4	365	7	BE818831	
C	46	15.4	367	4	BY674396	
47	15.4	90.6	383	4	BY674625	
C	48	15.4	410	7	AW206921	
C	49	15.4	449	7	BE818807	
50	15.4	90.6	451	14	DU607602	
51	15.4	90.6	494	3	BQ65719	
52	15.4	90.6	566	9	DA945234	
53	15.4	90.6	570	9	DA932540	
C	54	15.4	577	5	CK783403	
C	55	15.4	582	11	AZ259233	
C	56	15.4	586	12	CE755799	
C	57	15.4	593	9	DA940956	
58	15.4	90.6	600	5	CF751855	
59	15.4	90.6	601	8	CO196877	
60	15.4	90.6	609	13	CW480200	
61	15.4	90.6	669	2	BG293309	
C	62	15.4	671	13	CW480201	
C	63	15.4	689	9	DN265926	
64	15.4	90.6	694	9	DN441215	
65	15.4	90.6	721	8	CO196932	
C	66	15.4	744	12	CC111772	
67	15.4	90.6	749	13	CW576188	
C	68	15.4	788	8	CR374717	
C	69	15.4	797	7	BF700990	
C	70	15.4	798	8	CO196873	
C	71	15.4	802	11	BZ455483	
72	15.4	90.6	816	13	CW616903	
73	15.4	90.6	830	13	CW646415	
C	74	15.4	851	4	CA495187	
C	75	15.4	899	12	CC112512	
C	76	15.4	914	12	CC851671	
C	77	15.4	90.6	929	14	CNS02J9X
78	15.4	90.6	940	3	BQ685729	
79	15.4	90.6	1058	10	DM341470	
C	80	15.4	1117	9	DR125588	
C	81	15.4	1156	7	BE729821	
82	15.4	90.6	1263	5	CD256036	
83	15.4	90.6	1388	6	CNS0E2L3	
84	15.4	90.6	1454	6	CNS0G0ND	
85	15.4	90.6	1492	6	CNS0FYVC	
86	15.4	90.6	1505	6	CNS0G0M2S	
87	15.4	90.6	1512	6	CNS0G3DQ	
88	15.4	90.6	1518	6	CNS0EX00	
89	15.4	90.6	1962	14	DQ049847	
90	15.4	90.6	4208	6	BC068256	
91	15	88.2	198	2	BI529325	
C	92	15	278	1	AI613865	

93	15	88.2	309	10	DB992377	DR992377	Mdlr7011A	C 166	14.4	84.7	341	7	AW160798
94	15	88.2	322	7	AV850077	AV850077	AV850077	C 167	14.4	84.7	353	5	CI305825
95	15	88.2	368	4	CA561056	wlmi.pk00	CA561056	C 168	14.4	84.7	355	2	BM148829
96	15	88.2	419	3	BW421246	BW421246	BW421246	C 169	14.4	84.7	356	2	BI038890
97	15	88.2	438	8	CN899452	010706ABA	CN899452	C 170	14.4	84.7	364	4	C90338
98	15	88.2	484	3	BW366787	BW366787	BW366787	C 171	14.4	84.7	365	5	CK560001
99	15	88.2	492	3	BW178352	BW178352	BW178352	C 172	14.4	84.7	368	1	AA775070
100	15	88.2	517	3	BW203028	BW203028	BW203028	C 173	14.4	84.7	370	1	AI688285
101	15	88.2	518	7	AV884732	AV884732	AV884732	C 174	14.4	84.7	373	3	BQ738777
102	15	88.2	521	4	CA689031	wlms96.pk0	CA689031	C 175	14.4	84.7	379	1	AA994272
103	15	88.2	524	3	BW426366	BW426366	BW426366	C 176	14.4	84.7	384	8	CV127754
104	15	88.2	526	3	BW382890	BW382890	BW382890	C 177	14.4	84.7	392	2	BI534558
105	15	88.2	528	3	BW308736	BW308736	BW308736	C 178	14.4	84.7	392	8	CV095612
106	15	88.2	528	3	BW309783	BW309783	BW309783	C 179	14.4	84.7	404	3	BM839442
107	15	88.2	528	3	BW405257	BW405257	BW405257	C 180	14.4	84.7	409	1	AA159571
108	15	88.2	529	3	BW188031	BW188031	BW188031	C 181	14.4	84.7	414	5	CJ061166
109	15	88.2	534	8	CO903318	Mdb5022e	CO903318	C 182	14.4	84.7	429	7	BB875898
110	15	88.2	536	3	BW389352	BW389352	BW389352	C 183	14.4	84.7	431	8	CO872640
111	15	88.2	542	3	BW120319	BW120319	BW120319	C 184	14.4	84.7	445	10	N38821
112	15	88.2	543	3	BP024363	BP024363	BP024363	C 185	14.4	84.7	461	9	CX766672
113	15	88.2	554	3	BW472578	BW472578	BW472578	C 186	14.4	84.7	463	1	AA130015
114	15	88.2	554	7	AV854521	AV854521	AV854521	C 187	14.4	84.7	471	5	CI149721
115	15	88.2	554	7	AV976003	AV976003	AV976003	C 188	14.4	84.7	471	14	DU819050
116	15	88.2	557	3	BP019833	BP019833	BP019833	C 189	14.4	84.7	473	9	CX766470
117	15	88.2	560	3	BP019970	BP019970	BP019970	C 190	14.4	84.7	473	9	CX766497
118	15	88.2	561	3	BW306390	BW306390	BW306390	C 191	14.4	84.7	473	9	CX766740
119	15	88.2	565	3	BW095576	BW095576	BW095576	C 192	14.4	84.7	482	8	CV294645
120	15	88.2	565	7	AV962135	AV962135	AV962135	C 193	14.4	84.7	492	7	AW758506
121	15	88.2	566	7	AV864889	AV864889	AV864889	C 194	14.4	84.7	508	4	BX507570
122	15	88.2	567	3	BW315245	BW315245	BW315245	C 195	14.4	84.7	526	1	AA039719
123	15	88.2	571	3	BP015522	BP015522	BP015522	C 196	14.4	84.7	526	7	BF267413
124	15	88.2	572	3	BW41								









823	13.8	81.2	425	4	BY547337	BY547337	BY547337	C 896	13.8	81.2	454	9	DA552241	DA552241
824	13.8	81.2	425	4	BY633000	BY633000	BX633000	C 897	13.8	81.2	455	9	DA537667	DA537667
825	13.8	81.2	425	12	CC888292	SALK_1516	CC888292	C 898	13.8	81.2	456	3	BM836400	K-EST0112
826	13.8	81.2	426	4	BK475460	DKFZP686GJ	BK475460	C 899	13.8	81.2	457	9	DA083628	DA083628
827	13.8	81.2	426	7	BB728801	BB728801	BB728801	C 900	13.8	81.2	457	7	BE856791	7E68h02.X
828	13.8	81.2	428	9	DA994360	DA994360	DA994360	C 901	13.8	81.2	458	4	BY571085	BY571085
C 829	13.8	81.2	429	4	BY277616	BY277616	BY277616	C 902	13.8	81.2	458	9	DA987071	DA987071
C 830	13.8	81.2	429	4	BY288598	BY288598	BY288598	C 903	13.8	81.2	458	9	DB251063	DB251063
831	13.8	81.2	429	4	CY632109	CY632109	BY632109	C 904	13.8	81.2	459	7	BY599624	BY599624
C 832	13.8	81.2	429	5	CJ176724	CJ176724	CJ176724	C 905	13.8	81.2	459	7	AW465372	BP230019A
C 833	13.8	81.2	429	7	AW249026	AW249026	AW249026	C 906	13.8	81.2	459	7	DA087387	DA087387
C 834	13.8	81.2	430	5	CJ165752	CJ165752	CJ165752	C 907	13.8	81.2	460	4	CB737756	CB737756
C 835	13.8	81.2	431	1	AII155262	AII155262	uc87f06.f	C 908	13.8	81.2	460	5	CF339261	RCL1--04-
836	13.8	81.2	431	3	EM750982	EM750982	K-EST0026	C 909	13.8	81.2	460	9	DA8494	DA8494
837	13.8	81.2	431	4	BY593466	BY593466	BY593466	C 910	13.8	81.2	462	9	BI534940	BI534940
838	13.8	81.2	431	7	BB727333	BB727333	BB727333	C 911	13.8	81.2	463	14	CR912987	CR912987
839	13.8	81.2	431	14	AG976917	AG976917	Drosophila	C 912	13.8	81.2	463	4	BY251955	BY251955
C 840	13.8	81.2	432	4	BY167535	BY167535	BY167535	C 913	13.8	81.2	463	5	CF034830	OCG10d07.
C 841	13.8	81.2	432	4	BY211827	BY211827	BY211827	C 914	13.8	81.2	463	7	BE980981	UI-M-BH3-
842	13.8	81.2	432	4	BY593936	BY593936	BY593936	C 915	13.8	81.2	464	8	CV514070	TGESTzYr0
843	13.8	81.2	432	10	DV471104	DV471104	DV471104	C 916	13.8	81.2	466	4	BK432512	BK432512
C 844	13.8	81.2	432	10	T56251	T56251	YB32h01.s1	C 917	13.8	81.2	466	5	CF039208	OCH33d10.
845	13.8	81.2	433	8	CN199470	CN199470	TGESTzYl7	C 918	13.8	81.2	466	7	BB861165	BB861165
846	13.8	81.2	434	1	AA432216	AA432216	Vc35g12.f	C 919	13.8	81.2	466	8	CN336468	170004245
C 847	13.8	81.2	434	5	CJ183037	CJ183037	CJ183037	C 920	13.8	81.2	467	5	CF038757	QCH28C01.
848	13.8	81.2	434	7	BB788135	BB788135	BB788135	C 921	13.8	81.2	467	9	DA058037	DA058037
C 849	13.8	81.2	434	8	CN154135	CN154135	941322.MA	C 922	13.8	81.2	468	5	CF036902	QCQ38e11.
C 850	13.8	81.2	434	8	CV606283	CV606283	mdu94f11.	C 923	13.8	81.2	468	7	BF599892	BF599892
C 851	13.8	81.2	4											





derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

## ORIGIN

Query Match 100.0%; Score 17; DB 7; Length 307;  
 Best Local Similarity 100.0%; Pred. No. 5.3e+02;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCCGAG 17  
 |||||  
 Db 242 GCCCAGCTTGGCCGAG 226

## RESULT 3

BE818817/c  
 LOCUS  
 DEFINITION CM2-BN0302-050700-256-b07 BN0302 Homo sapiens cDNA, mRNA sequence. EST 21-SEP-2000  
 ACCESSION BE818817  
 VERSION BE818817.1 GI:10251051  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 341)  
 AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.P., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,P., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 PUBMED 10737800  
 COMMENT Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2=CM2-BN0302-050700-256-b07&t3=2000-07-05&t4=1)

Seq primer: puc 18 forward  
 High quality sequence start: 10  
 High quality sequence stop: 341.

## FEATURES

source  
 1..341  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /dev\_stage="Adult"  
 /clone\_lib="BN0302"  
 /note="Organ: breast normal; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

## ORIGIN

Query Match 100.0%; Score 17; DB 7; Length 341;  
 Best Local Similarity 100.0%; Pred. No. 5.3e+02;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCCGAG 17  
 |||||  
 Db 253 GCCCAGCTTGGCCGAG 237

## RESULT 4

AA112374  
 LOCUS  
 DEFINITION z68e10.r1 Stratagene HeLa cell s3 937216 Homo sapiens cDNA clone IMAGE:563370 5', mRNA sequence.

ACCESSION AA112374  
 VERSION AA112374.1 GI:1664784  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 367)  
 AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B., Chisoe,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W., Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevisakis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.

Generation and analysis of 280,000 human expressed sequence tags  
 Genome Res. 6 (9), 807-828 (1996)  
 8889549  
 CONTACT: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
 Insert Length: 1363 Std Error: 0.00  
 Seq primer: -28M13 rev2 from Amersham  
 High quality sequence stop: 329.  
 Location/Qualifiers  
 1..367  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="GDB:4593181"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:563370"  
 /sex="female"  
 /dev\_stage="HeLa S3 cell line"  
 /clone\_lib="SOLR (kanamycin resistant)"  
 /note="vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer: Oligo dt. HeLa S3 epitheloid carcinoma cells grown to semi-confluency without induction. Average insert size: 1.5 kb; Uni-ZAP XR Vector. -5' adaptor sequence: 5' GAATTCGCGCAG 3' -3' adaptor sequence: 5' CTCGATTTTTTTTTTTTTTTT 3"

## FEATURES

source  
 1..367  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="GDB:4593181"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:563370"  
 /sex="female"  
 /dev\_stage="HeLa S3 cell line"  
 /clone\_lib="SOLR (kanamycin resistant)"  
 /note="vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer: Oligo dt. HeLa S3 epitheloid carcinoma cells grown to semi-confluency without induction. Average insert size: 1.5 kb; Uni-ZAP XR Vector. -5' adaptor sequence: 5' GAATTCGCGCAG 3' -3' adaptor sequence: 5' CTCGATTTTTTTTTTTTTTTT 3"

## ORIGIN

Query Match 100.0%; Score 17; DB 1; Length 367;  
 Best Local Similarity 100.0%; Pred. No. 5.4e+02;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCCGAG 17  
 |||||  
 Db 116 GCCCAGCTTGGCCGAG 132

## RESULT 5

BE818835/c  
 LOCUS  
 DEFINITION CM2-BN0302-050700-256-cl1 BN0302 Homo sapiens cDNA, mRNA sequence.

```

ACCESSION      BE818835
VERSION        BE818835.1  GI:10251069
KEYWORDS       EST.
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
               Homnidae; Homo.
REFERENCE      1 (bases 1 to 384)
AUTHORS        Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
               Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
               Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
               O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
               Simpson,A.J.
TITLE          Shotgun sequencing of the human transcriptome with ORF expressed
               sequence tags
JOURNAL        Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
PUBMED        10737800
COMMENT        Contact: Simpson A.J.G.
               Laboratory of Cancer Genetics
               Ludwig Institute for Cancer Research
               Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
               Brazil
               Tel: +55-11-2704922
               Fax: +55-11-2707001
               Email: asimpson@ludwig.org.br
               This sequence was derived from the FAPESP/LICR Human Cancer Genome
               Project. This entry can be seen in the following URL
               (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2=CM2-BN0302-050
               700-256-a05&tl3=2000-07-05&tl4=1)
               Seq primer: puc 18 forward
               High quality sequence start: 5
               High quality sequence stop: 384.
FEATURES       source
               1..384
               Location/Qualifiers
               1..396
               /organism="Homo sapiens"
               /mol_type="mRNA"
               /db_xref="taxon:9606"
               /dev_stage="Adult"
               /clone_lib="BN0302"
               /notes="Organ: breast normal; Vector: puc18; Site_1: SmaI;
               Site_2: SmaI; A mini-library was made by cloning products
               derived from ORESTES PCR (U.S. Letters Patent application
               No. 196,716 - Ludwig Institute for Cancer Research)
               profiles into the pUC 18 vector. Reverse transcription of
               tissue mRNA and cDNA amplification were performed under
               low stringency conditions."
ORIGIN
Query Match      100.0%; Score 17; DB 7; Length 384;
Best Local Similarity 100.0%; Pred. No. 5.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17
   |||||||||||||||
Db 257 GCCCAGCGTTGGCCGAG 241

RESULT 6
BE818828/c
LOCUS          BE818828
DEFINITION    CM2-BN0302-050700-256-a05 BN0302 Homo sapiens cDNA, mRNA sequence.
ACCESSION     BE818828
VERSION       BE818828.1  GI:10251062
KEYWORDS       EST.
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
               Homnidae; Homo.
REFERENCE      1 (bases 1 to 396)
AUTHORS        Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
               Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
               Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
               Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
               O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
               Simpson,A.J.
TITLE          Shotgun sequencing of the human transcriptome with ORF expressed
               sequence tags
JOURNAL        Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
PUBMED        10737800
COMMENT        Contact: Simpson A.J.G.
               Laboratory of Cancer Genetics
               Ludwig Institute for Cancer Research
               Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
               Brazil
               Tel: +55-11-2704922
               Fax: +55-11-2707001
               Email: asimpson@ludwig.org.br
               This sequence was derived from the FAPESP/LICR Human Cancer Genome
               Project. This entry can be seen in the following URL
               (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2=CM2-BN0302-050
               700-256-a05&tl3=2000-07-05&tl4=1)
               Seq primer: puc 18 forward
               High quality sequence start: 5
               High quality sequence stop: 384.
FEATURES       source
               1..384
               Location/Qualifiers
               1..396
               /organism="Homo sapiens"
               /mol_type="mRNA"
               /db_xref="taxon:9606"
               /dev_stage="Adult"
               /clone_lib="BN0302"
               /notes="Organ: breast normal; Vector: puc18; Site_1: SmaI;
               Site_2: SmaI; A mini-library was made by cloning products
               derived from ORESTES PCR (U.S. Letters Patent application
               No. 196,716 - Ludwig Institute for Cancer Research)
               profiles into the pUC 18 vector. Reverse transcription of
               tissue mRNA and cDNA amplification were performed under
               low stringency conditions."
ORIGIN
Query Match      100.0%; Score 17; DB 7; Length 384;
Best Local Similarity 100.0%; Pred. No. 5.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17
   |||||||||||||||
Db 257 GCCCAGCGTTGGCCGAG 241

RESULT 7
BE818840/c
LOCUS          BE818840
DEFINITION    CM2-BN0302-050700-256-f02 BN0302 Homo sapiens cDNA, mRNA sequence.
ACCESSION     BE818840
VERSION       BE818840.1  GI:10251074
KEYWORDS       EST.
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
               Homnidae; Homo.
REFERENCE      1 (bases 1 to 404)
AUTHORS        Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
               Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
               Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
               Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
               O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
               Simpson,A.J.
TITLE          Shotgun sequencing of the human transcriptome with ORF expressed
               sequence tags
JOURNAL        Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
PUBMED        10737800
COMMENT        Contact: Simpson A.J.G.

```

Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome  
 Project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-CM2-BN0302-050  
 700-256-f02&t3=2000-07-05&t4=1)

Seq primer: puc 18 forward  
 High quality sequence start: 27  
 High quality sequence stop: 404.

## FEATURES

Location/Qualifiers  
 source  
 1..404  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /dev\_stage="Adult"  
 /clone\_lib="BN0302"

/note="Organ: breast normal; Vector: puc18; Site: 1: SmaI;  
 Site 2: SmaI; A mini-library was made by cloning products  
 derived from ORESTES PCR (U.S. Letters Patent application  
 No. 196,716 - Ludwig Institute for Cancer Research)  
 profiles into the pUC 18 vector. Reverse transcription of  
 tissue mRNA and cDNA amplification were performed under  
 low stringency conditions."

## ORIGIN

Query Match 100.0%; Score 17; DB 7; Length 404;  
 Best Local Similarity 100.0%; Pred. No. 5.4e+02;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17

Db 244 GCCCAGCGTTGGCCGAG 228

## RESULT 8

BE818805/c  
 LOCUS BE818805 424 bp mRNA linear EST 21-SEP-2000  
 DEFINITION CM2-BN0302-040700-253-b10 BN0302 Homo sapiens cDNA, mRNA sequence.

ACCESSION BE818805

VERSION BE818805.1 GI:10251039

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homiidae; Homo.

REFERENCE 1 (bases 1 to 424)

Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
 Goldman, G.H., Carvalho, A.P., Matsukuma, A., Baia, G.S., Simpson, D.H.,  
 Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V.,  
 O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
 Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

10737800

## COMMENT

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome  
 Project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-CM2-BN0302-040

700-253-b10&t3=2000-07-04&t4=1)  
 Seq primer: puc 18 forward  
 High quality sequence start: 4  
 High quality sequence stop: 424.

## FEATURES

Location/Qualifiers  
 source  
 1..424  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /dev\_stage="Adult"  
 /clone\_lib="BN0302"  
 /note="Organ: breast normal; Vector: puc18; Site: 1: SmaI;  
 Site 2: SmaI; A mini-library was made by cloning products  
 derived from ORESTES PCR (U.S. Letters Patent application  
 No. 196,716 - Ludwig Institute for Cancer Research)  
 profiles into the pUC 18 vector. Reverse transcription of  
 tissue mRNA and cDNA amplification were performed under  
 low stringency conditions."

## ORIGIN

Query Match 100.0%; Score 17; DB 7; Length 424;  
 Best Local Similarity 100.0%; Pred. No. 5.4e+02;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17

Db 250 GCCCAGCGTTGGCCGAG 234

## RESULT 9

CD672930  
 LOCUS fg18a05.y2 Human Iris cDNA (Normalized): fg Homo sapiens cDNA clone  
 DEFINITION fg18a05.5', mRNA sequence.

ACCESSION CD672930

VERSION CD672930.1 GI:32174661

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homiidae; Homo.

REFERENCE 1 (bases 1 to 453)

Wistow, G., Bernstein, S.L., Ray, S., Wyatt, M.K., Behal, A.,

Touchman, J.W., Bouffard, G., Smith, D. and Peterson, K.

Expressed sequence tag analysis of adult human iris for the NEtBank  
 Project: steroid-response factors and similarities with retinal  
 pigment epithelium

Mol. Vis. 8 (4), 185-195 (2002)

12107412

Contact: Wistow G

Section on Molecular Structure and Function

National Eye Institute

6/331, NIH, Bethesda, MD 20892-2740, USA

Tel: 301 402 3452

Fax: 301 496 0078

Email: graeme@helix.nih.gov

Plate: 18 row: a column: 05

Seq primer: M13RPI reverse primer (ABI).

## FEATURES

Location/Qualifiers  
 source  
 1..453  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone\_lib="fg18a05"  
 /tissue\_type="Iris"  
 /dev\_stage="Adult"  
 /lab\_host="EMDH10B"

/clone\_lib="Human Iris cDNA (Normalized): fg"

/note="Organ: Eye; Vector: pCMVSPORT6; A human iris  
 library (bx) was normalized by self-subtraction. One

portion of double stranded plasmid DNA representing the  
 library was linearized by NotI. This NotI digested library

was used as a template for biotinylated RNA synthesis using SP6 RNA polymerase. Another portion of the double stranded plasmid library was converted to single-stranded circles in vitro using Gene II and Exonuclease III (Life Technologies). Single-stranded DNA (1 mg) was hybridized (Cot 500) with 41 mg of Bio-RNA and vector blocking oligonucleotides. The hybridized Bio-RNA/ss-circles were removed by streptavidin:phenol extraction. EST analysis was performed on the library at the NIH Intramural Sequencing Center (NISC)."

## ORIGIN

Query Match 100.0%; Score 17; DB 5; Length 453;  
 Best Local Similarity 100.0%; Pred. No. 5.4e+02;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGCCGAG 17  
 |||||  
 Db 56 GCCACGCTTGCCGAG 72

## RESULT 10

BE818871/c  
 LOCUS BE818871 458 bp mRNA linear EST 21-SEP-2000  
 DEFINITION CM2-BN0302-100700-259-f01 BN0302 Homo sapiens cDNA, mRNA sequence.  
 ACCESSION BE818871  
 VERSION BE818871.1 GI:10251105  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo.

1 (bases 1 to 458)

Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.  
 Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL  
 PUBLISHED Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 10737800

COMMENT Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001

Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2=CM2-BN0302-100700-259-f01&t3=2000-07-10&t4=1)  
 Seq primer: puc 18 forward  
 High quality sequence start: 12  
 High quality sequence stop: 458.

## FEATURES

## source

1. 458  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /dev\_stage="adult"  
 /clone\_lib="BN0302"  
 /note="Organ: breast\_normal; Vector: puc18; Site:1: SmaI; Site:2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

## ORIGIN

Query Match 100.0%; Score 17; DB 7; Length 458;  
 Best Local Similarity 100.0%; Pred. No. 5.4e+02;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGCCGAG 17  
 |||||  
 Db 264 GCCACGCTTGCCGAG 248

## RESULT 11

BE818859/c  
 LOCUS BE818859 466 bp mRNA linear EST 21-SEP-2000  
 DEFINITION CM2-BN0302-100700-259-c02 BN0302 Homo sapiens cDNA, mRNA sequence.  
 ACCESSION BE818859  
 VERSION BE818859.1 GI:10251093  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo.

1 (bases 1 to 466)

Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.  
 Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL  
 PUBLISHED Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 10737800

COMMENT Contact: Simpson A.J.G.

Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001

Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2=CM2-BN0302-100700-259-c02&t3=2000-07-10&t4=1)  
 Seq primer: puc 18 forward  
 High quality sequence start: 16  
 High quality sequence stop: 466.

## FEATURES

## source

1. 466  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /dev\_stage="Adult"  
 /clone\_lib="BN0302"  
 /note="Organ: breast\_normal; Vector: puc18; Site:1: SmaI; Site:2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

## ORIGIN

Query Match 100.0%; Score 17; DB 7; Length 466;  
 Best Local Similarity 100.0%; Pred. No. 5.4e+02;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGCCGAG 17  
 |||||  
 Db 267 GCCACGCTTGCCGAG 251

```

RESULT 12
BE818852/c
LOCUS BE818852 467 bp mRNA linear EST 21-SEP-2000
DEFINITION CM3-BN0302-100700-259-a05 BN0302 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE818852
VERSION BE818852.1 GI:10251086
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 467)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baid,G.S., Simpson,D.H.,
Brenstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V.,
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
PUBMED 10737800
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?ti=CM2-BN0302-100
700-259-a05&t3=2000-07-10&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 19
High quality sequence stop: 467.
FEATURES
Location/Qualifiers
source
1..467
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="BN0302"
/note="Organ: breast_normal; Vector: puc18; Site 1: SmaI;
Site 2: SmaI; A mini-library was made by cloning products
derived from ORFESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
ORIGIN
Query Match 100.0%; Score 17; DB 7; Length 467;
Best Local Similarity 100.0%; Pred. No. 5.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 GCCCAGCGTGGCCGAG 17
|||||
Db 246 GCCCAGCGTGGCCGAG 230
|||||

RESULT 13
AW390096
LOCUS AW390096 528 bp mRNA linear EST 04-FEB-2000
DEFINITION CM3-ST0179-051099-019-f07 ST0179 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW390096
VERSION AW390096.1 GI:6894755
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 528)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baid,G.S., Simpson,D.H.,
Brenstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V.,
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
PUBMED 10737800
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?ti=CM2-BN0302-100
700-259-a05&t3=2000-07-10&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 19
High quality sequence stop: 467.
FEATURES
Location/Qualifiers
source
1..528
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="ST0179"
/note="Organ: stomach; Vector: puc18; Site 1: SmaI;
Site 2: SmaI; A mini-library was made by cloning products
derived from ORFESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
ORIGIN
Query Match 100.0%; Score 17; DB 7; Length 528;
Best Local Similarity 100.0%; Pred. No. 5.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 GCCCAGCGTGGCCGAG 17
|||||
Db 502 GCCCAGCGTGGCCGAG 518
|||||

RESULT 14
BP253378
LOCUS BP253378 571 bp mRNA linear EST 15-SEP-2004
DEFINITION BP253378 Sugano cDNA library, kidney epithelial cell Homo sapiens
cDNA clone HRC03720, mRNA sequence.
ACCESSION BP253378
VERSION BP253378.1 GI:52135659
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 571)
Suzuki,Y., Yamashita,R., Shiota,M., Sakakibara,Y., Chiba,J.,
Mizushima-Sugano,J., Nakai,K. and Sugano,S.
Sequence comparison of human and mouse genes reveals a homologous
block structure in the promoter regions
Genome Res. 14 (9), 1711-1718 (2004)
15342556
Contact: Yutaka Suzuki
Department of Virology
Institute of Medical Science, University of Tokyo
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
Tel: 81-3-5449-5343
Fax: 81-3-5449-5416

```

```

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo
1 (bases 1 to 528)
HCGP http://www.ludwig.org.br/ORESTES.
The FAPESP/LICR Human Cancer Genome Project
Unpublished (1999)
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?ti=CM3&t2=CM3-ST0179-
051099-019-f07&t3=1999-10-05&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 18
High quality sequence stop: 528.
FEATURES
Location/Qualifiers
source
1..528
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="ST0179"
/note="Organ: stomach; Vector: puc18; Site 1: SmaI;
Site 2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
ORIGIN
Query Match 100.0%; Score 17; DB 7; Length 528;
Best Local Similarity 100.0%; Pred. No. 5.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 GCCCAGCGTGGCCGAG 17
|||||
Db 502 GCCCAGCGTGGCCGAG 518
|||||

RESULT 14
BP253378
LOCUS BP253378 571 bp mRNA linear EST 15-SEP-2004
DEFINITION BP253378 Sugano cDNA library, kidney epithelial cell Homo sapiens
cDNA clone HRC03720, mRNA sequence.
ACCESSION BP253378
VERSION BP253378.1 GI:52135659
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 571)
Suzuki,Y., Yamashita,R., Shiota,M., Sakakibara,Y., Chiba,J.,
Mizushima-Sugano,J., Nakai,K. and Sugano,S.
Sequence comparison of human and mouse genes reveals a homologous
block structure in the promoter regions
Genome Res. 14 (9), 1711-1718 (2004)
15342556
Contact: Yutaka Suzuki
Department of Virology
Institute of Medical Science, University of Tokyo
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
Tel: 81-3-5449-5343
Fax: 81-3-5449-5416

```

```

Email: yeuzuki@hgc.jp.
Location/Qualifiers
source
  1. 571
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone="HRC03720"
    /tissue_type="kidney"
    /cell_type="epithelial cell"
    /clone_lib="Sugano cDNA library, kidney epithelial cell"

ORIGIN
Query Match      100.0%; Score 17; DB 3; Length 571;
Best Local Similarity 100.0%; Pred. No. 5.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCGGAG 17
Db 522 GCCCAGCGTTGGCGGAG 538

RESULT 15
CB152636      577 bp mRNA linear EST 29-JAN-2003
LOCUS      K-EST0209868 L12JSHC0e1 Homo sapiens cDNA clone L12JSHC0e1-5-G07
DEFINITION      5', mRNA sequence.
ACCESSION      CB152636
VERSION      CB152636.1 GI:28137590
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
               Hominidae; Homo.
REFERENCE      1. (bases 1 to 577)
               Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
               Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
               Kim,Y.S.
               21C Frontier Korean EST Project 2001
               Unpublished (2002)
               Contact: Kim YS
               Genome Research Center
               Korea Research Institute of Bioscience & Biotechnology
               52 Eosun-dong Yuseong-gu, Daejeon 305-333, South Korea
               Tel: +82-42-860-4470
               Fax: +82-42-860-4409
               Email: vongsung@mail.kribb.re.kr
               Plate: 5 row: G column: 07
               High quality sequence stop: 577.
               Location/Qualifiers
               1. 577
                 /organism="Homo sapiens"
                 /mol_type="mRNA"
                 /db_xref="taxon:9606"
                 /clone="L12JSHC0e1-5-G07"
                 /sex="M"
                 /cell_line="J-SHC"
                 /lab_host="Top10F"
                 /clone_lib="L12JSHC0e1"
                 /note="Organ: Liver; Vector: pcNS-D2, Site 1: EcoRI;
                 Site 2: NotI; The poly (A)+ RNA was dephosphorylated with
                 bacterial alkaline phosphatase (BAP) and then decapped
                 with tabacco acid pyrophosphatase (TAP). The decapped
                 intact mRNA was ligated with DNA-RNA linker including
                 EcoRI site by treatment of T4 RNA ligase and the first
                 strand cDNA was synthesized from oligo dT-selected mRNA by
                 priming with dT-tailed vector. The dT-tailed vector was
                 adjusted to have about 60nt. The cDNA vector was
                 circularized with E. coli DNA ligase after digestion of
                 EcoRI which site is also included in vector. An RNA strand
                 converted to a DNA strand by Okayama-Berg method. The
                 obtained cDNA vectors were used for transformation of
                 competent cells E. coli Top10F by electroporation method.

FEATURES
source
  1. 577
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone="L12JSHC0e1-5-G07"
    /sex="M"
    /cell_line="J-SHC"
    /lab_host="Top10F"
    /clone_lib="L12JSHC0e1"
    /note="Organ: Liver; Vector: pcNS-D2, Site 1: EcoRI;
    Site 2: NotI; The poly (A)+ RNA was dephosphorylated with
    bacterial alkaline phosphatase (BAP) and then decapped
    with tabacco acid pyrophosphatase (TAP). The decapped
    intact mRNA was ligated with DNA-RNA linker including
    EcoRI site by treatment of T4 RNA ligase and the first
    strand cDNA was synthesized from oligo dT-selected mRNA by
    priming with dT-tailed vector. The dT-tailed vector was
    adjusted to have about 60nt. The cDNA vector was
    circularized with E. coli DNA ligase after digestion of
    EcoRI which site is also included in vector. An RNA strand
    converted to a DNA strand by Okayama-Berg method. The
    obtained cDNA vectors were used for transformation of
    competent cells E. coli Top10F by electroporation method.

The cDNA libraries constructed by this method are
full-length enriched cDNA library. After analyzing and
sequencing about 2,000 - 3,000 colonies in original cDNA
library, the abundant cDNAs were selected and amplified by
PCR reaction using vector region primer including T7
promotor as 5' primer and N(dT)14 as 3' primer. The PCR
products were used as template for synthesis of
biotinylated single stranded RNA by in vitro transcription
reaction. The synthesized RNA probes were hybridized with
antisense single stranded cDNAs prepared from original
library and incubated with avidin-gel. After removing
DNA-RNA hybrids by centrifuge, the substracted cDNA
libraries were constructed by transformatoin of the
remaining DNA into competent cells E. coli Top10F with
electroporation method."
```

```

The cDNA libraries constructed by this method are
full-length enriched cDNA library. After analyzing and
sequencing about 2,000 - 3,000 colonies in original cDNA
library, the abundant cDNAs were selected and amplified by
PCR reaction using vector region primer including T7
promotor as 5' primer and N(dT)14 as 3' primer. The PCR
products were used as template for synthesis of
biotinylated single stranded RNA by in vitro transcription
reaction. The synthesized RNA probes were hybridized with
antisense single stranded cDNAs prepared from original
library and incubated with avidin-gel. After removing
DNA-RNA hybrids by centrifuge, the substracted cDNA
libraries were constructed by transformatoin of the
remaining DNA into competent cells E. coli Top10F with
electroporation method."
```

ORIGIN

Query Match 100.0%; Score 17; DB 4; Length 577;  
Best Local Similarity 100.0%; Pred. No. 5.4e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCGGAG 17  
Db 197 GCCCAGCGTTGGCGGAG 213

RESULT 16  
BP267556 578 bp mRNA linear EST 16-SEP-2004  
LOCUS BP267556 Sugano cDNA library, thyroid JTH Homo sapiens cDNA clone  
DEFINITION JTH07367, mRNA sequence.  
ACCESSION BP267556  
VERSION BP267556.1 GI:52182788  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Hominidae; Homo.  
REFERENCE 1. (bases 1 to 578)  
 Suzuki,Y., Yamashita,R., Shiota,M., Sakakibara,Y., Chiba,J.,  
 Mizushima-Sugano,J., Nakai,K. and Sugano,S.  
 Sequence comparison of human and mouse genes reveals a homologous  
 block structure in the promoter regions  
 Genome Res. 14 (9), 1711-1718 (2004)  
 15342556  
 Contact: Yutaka Suzuki  
 Department of Virology  
 Institute of Medical Science, University of Tokyo  
 4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
 Tel: 81-3-5449-5343  
 Fax: 81-3-5449-5416  
 Email: ysuzuki@hgc.jp.  
 Location/Qualifiers  
 1. 578  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="JTH07367"  
 /tissue\_type="thyroid"  
 /cell\_line="JTH"  
 /clone\_lib="Sugano cDNA library, thyroid JTH"  
 /note="thyroid tumor"

ORIGIN

Query Match 100.0%; Score 17; DB 3; Length 578;  
Best Local Similarity 100.0%; Pred. No. 5.4e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCGGAG 17  
Db 556 GCCCAGCGTTGGCGGAG 572

```
RESULT 17
BP265234
LOCUS
DEFINITION BP265234 Sugano cDNA library, thyroid JTH Homo sapiens cDNA clone
JTH00322, mRNA sequence.
ACCESSION
VERSION BP265234.1 GI:52180465
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1 (bases 1 to 580)
AUTHORS Suzuki, Y., Yamashita, R., Shirota, M., Sakakibara, Y., Chiba, J.,
Mizushima-Sugano, J., Nakai, K. and Sugano, S.
TITLE Sequence comparison of human and mouse genes reveals a homologous
block structure in the promoter regions
JOURNAL Genome Res. 14 (9), 1711-1718 (2004)
PUBMED 15342556
COMMENT Contact: Yutaka Suzuki
Department of Virology
Institute of Medical Science, University of Tokyo
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
Tel: 81-3-5449-5343
Fax: 81-3-5449-5416
Email: yuzuki@hgc.jp.
FEATURES
source
1. 580
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="JTH00322"
/tissue_type="thyroid"
/cell_line="JTH"
/clone_lib="Sugano cDNA library, thyroid JTH"
/note="thyroid tumor"
ORIGIN
Query Match 100.0%; Score 17; DB 3; Length 580;
Best Local Similarity 100.0%; Pred. No. 5.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GCCCAGCGTTGGCCGAG 17
|||||
Db 327 GCCCAGCGTTGGCCGAG 343

RESULT 18
BP267627
LOCUS
DEFINITION BP267627 Sugano cDNA library, thyroid JTH Homo sapiens cDNA clone
JTH07584, mRNA sequence.
ACCESSION
VERSION BP267627.1 GI:52182859
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1 (bases 1 to 580)
AUTHORS Suzuki, Y., Yamashita, R., Shirota, M., Sakakibara, Y., Chiba, J.,
Mizushima-Sugano, J., Nakai, K. and Sugano, S.
TITLE Sequence comparison of human and mouse genes reveals a homologous
block structure in the promoter regions
JOURNAL Genome Res. 14 (9), 1711-1718 (2004)
PUBMED 15342556
COMMENT Contact: Yutaka Suzuki
Department of Virology
Institute of Medical Science, University of Tokyo
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
```

```
Tel: 81-3-5449-5343
Fax: 81-3-5449-5416
Email: yuzuki@hgc.jp.
FEATURES
source
1. 580
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="JTH07584"
/tissue_type="thyroid"
/cell_line="JTH"
/clone_lib="Sugano cDNA library, thyroid JTH"
/note="thyroid tumor"
ORIGIN
Query Match 100.0%; Score 17; DB 3; Length 580;
Best Local Similarity 100.0%; Pred. No. 5.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GCCCAGCGTTGGCCGAG 17
|||||
Db 539 GCCCAGCGTTGGCCGAG 555

RESULT 19
BP268151
LOCUS
DEFINITION BP268151 Sugano cDNA library, thyroid JTH Homo sapiens cDNA clone
JTH09144, mRNA sequence.
ACCESSION
VERSION BP268151.1 GI:52183383
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1 (bases 1 to 580)
AUTHORS Suzuki, Y., Yamashita, R., Shirota, M., Sakakibara, Y., Chiba, J.,
Mizushima-Sugano, J., Nakai, K. and Sugano, S.
TITLE Sequence comparison of human and mouse genes reveals a homologous
block structure in the promoter regions
JOURNAL Genome Res. 14 (9), 1711-1718 (2004)
PUBMED 15342556
COMMENT Contact: Yutaka Suzuki
Department of Virology
Institute of Medical Science, University of Tokyo
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
Tel: 81-3-5449-5343
Fax: 81-3-5449-5416
Email: yuzuki@hgc.jp.
FEATURES
source
1. 580
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="JTH09144"
/tissue_type="thyroid"
/cell_line="JTH"
/clone_lib="Sugano cDNA library, thyroid JTH"
/note="thyroid tumor"
ORIGIN
Query Match 100.0%; Score 17; DB 3; Length 580;
Best Local Similarity 100.0%; Pred. No. 5.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GCCCAGCGTTGGCCGAG 17
|||||
Db 538 GCCCAGCGTTGGCCGAG 554

RESULT 20
```

```

BP256144
LOCUS      BP256144      582 bp      mRNA      linear      EST 16-SEP-2004
DEFINITION BP256144 Sugano cDNA library, kidney epithelial cell Homo sapiens
            cDNA clone HRC10671, mRNA sequence.
ACCESSION  BP256144
VERSION    BP256144.1 GI:52171374
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homnidae; Homo.
REFERENCE  1 (bases 1 to 582)
AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
PUBMED    15342556
COMMENT   Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Tel: 81-3-5449-5343
            Fax: 81-3-5449-5416
            Email: yusuzuki@hgc.jp.

FEATURES             source
            1..582
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="HRC10671"
            /tissue_type="kidney"
            /cell_type="epithelial cell"
            /clone_lib="Sugano cDNA library, kidney epithelial cell"

ORIGIN
Query Match      100.0%; Score 17; DB 3; Length 582;
Best Local Similarity 100.0%; Pred. No. 5.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1  GCCACGCTTGGCCGAG 17
        |||||||
Db      415 GCCACGCTTGGCCGAG 431

RESULT 21
BP267918
LOCUS      BP267918      582 bp      mRNA      linear      EST 16-SEP-2004
DEFINITION BP267918 Sugano cDNA library, thyroid JTH Homo sapiens cDNA clone
            JTH08419, mRNA sequence.
ACCESSION  BP267918
VERSION    BP267918.1 GI:52183150
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homnidae; Hmo.
REFERENCE  1 (bases 1 to 582)
AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
PUBMED    15342556
COMMENT   Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Tel: 81-3-5449-5343
            Fax: 81-3-5449-5416
            Email: yusuzuki@hgc.jp.

FEATURES             source
            1..582
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="JTH08419"
            /tissue_type="thyroid"
            /cell_line="JTH"
            /clone_lib="Sugano cDNA library, thyroid JTH"
            /note="thyroid tumor"

ORIGIN
Query Match      100.0%; Score 17; DB 3; Length 582;
Best Local Similarity 100.0%; Pred. No. 5.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1  GCCACGCTTGGCCGAG 17
        |||||||
Db      556 GCCACGCTTGGCCGAG 572

RESULT 22
BP363758
LOCUS      BP363758      582 bp      mRNA      linear      EST 17-SEP-2004
DEFINITION BP363758 Sugano cDNA library, fetal lung fibroblast TIG Homo
            sapiens cDNA clone TIR02383, mRNA sequence.
ACCESSION  BP363758
VERSION    BP363758.1 GI:52293963
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homnidae; Homo.
REFERENCE  1 (bases 1 to 582)
AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
PUBMED    15342556
COMMENT   Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Tel: 81-3-5449-5343
            Fax: 81-3-5449-5416
            Email: yusuzuki@hgc.jp.

FEATURES             source
            1..582
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="TIR02383"
            /tissue_type="lung"
            /cell_type="fibroblast"
            /cell_line="TIG"
            /dev_stage="fetal"
            /clone_lib="Sugano cDNA library, fetal lung fibroblast
            TIG"

ORIGIN
Query Match      100.0%; Score 17; DB 3; Length 582;
Best Local Similarity 100.0%; Pred. No. 5.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1  GCCACGCTTGGCCGAG 17
        |||||||
Db      380 GCCACGCTTGGCCGAG 396

RESULT 23
BE006062

```



```

LOCUS       BE006062               583 bp      mRNA      linear      EST 05-JUN-2000
DEFINITION   RC2-BN0123-170300-011-c07 BN0123 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE006062
VERSION     BE006062.1   GI:8266295
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1 (bases 1 to 583)
AUTHORS     Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.P.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V.,
            O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
TITLE       Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
JOURNAL     Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
PUBMED     10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2-RC2-BN0123-170
            300-011-c07&t3=2000-03-17&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 24
High quality sequence stop: 583.
            Location/Qualifiers
                source
                    1..583
                        /organism="Homo sapiens"
                        /mol_type="mRNA"
                        /db_xref="taxon:9606"
                        /dev_stage="Adult"
                        /clone_lib="BN0123"
                        /note="Organ: breast normal; Vector: puc18; Site 1: SnaI;
                        Site 2: SnaI; A mini-library was made by cloning products
                        derived from ORESTES PCR (U.S. Letters Patent application
                        No. 196,716 - Ludwig Institute for Cancer Research)
                        profiles into the puc 18 vector. Reverse transcription of
                        tissue mRNA and cDNA amplification were performed under
                        low stringency conditions."

FEATURES             source
Query Match       100.0%; Score 17; DB 7; Length 583;
Best Local Similarity 100.0%; Pred. No. 5.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy       1   GCCCAGCTTGGCCGAG 17
          |||||
Db       552 GCCCAGCTTGGCCGAG 568

RESULT 24
LOCUS    BP256147               584 bp      mRNA      linear      EST 16-SEP-2004
DEFINITION   BP256147 Sugano cDNA library, kidney epithelial cell Homo sapiens
            cDNA clone HRC10676, mRNA sequence.
ACCESSION   BP256147
VERSION     BP256147.1   GI:52171377
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1 (bases 1 to 584)
AUTHORS     Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
TITLE       Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
JOURNAL     Genome Res. 14 (9), 1711-1718 (2004)
PUBMED     15342556
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Tel: 81-3-5449-5343
            Fax: 81-3-5449-5416
            Email: yezukui@bmc.jp.
            Location/Qualifiers
                source
                    1..584
                        /organism="Homo sapiens"
                        /mol_type="mRNA"
                        /db_xref="taxon:9606"
                        /clone="HRC10676"
                        /tissue_type="kidney"
                        /cell_type="epithelial cell"
                        /clone_lib="Sugano cDNA library, kidney epithelial cell"

ORIGIN
Query Match       100.0%; Score 17; DB 3; Length 584;
Best Local Similarity 100.0%; Pred. No. 5.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy       1   GCCCAGCTTGGCCGAG 17
          |||||
Db       416 GCCCAGCTTGGCCGAG 432

RESULT 25
LOCUS    CB130950               621 bp      mRNA      linear      EST 29-JAN-2003
DEFINITION   K-EST0180946 L12JSHCO Homo sapiens cDNA clone L12JSHCO-5-D03 5',
            mRNA sequence.
ACCESSION   CB130950
VERSION     CB130950.1   GI:28095412
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1 (bases 1 to 621)
AUTHORS     Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
            Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
            Kim,Y.S.
TITLE       21C Frontier Korean EST Project 2001
JOURNAL     Unpublished (2002)
COMMENT     Contact: Kim YS
            Genome Research Center
            Korea Research Institute of Bioscience & Biotechnology
            52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
            Tel: +82-42-860-4470
            Fax: +82-42-860-4409
            Email: yongsung@mail.kribb.re.kr
            Plate: 5 row: D column: 03
            High quality sequence stop: 621.
            Location/Qualifiers
                source
                    1..621
                        /organism="Homo sapiens"
                        /mol_type="mRNA"
                        /db_xref="taxon:9606"
                        /clone="L12JSHCO-5-D03"
                        /sex="M"
                        /cell_line="J-SHC"
                        /lab_host="Top10F"
                        /clone_lib="L12JSHCO"

```

/note="Organ: Liver; Vector: pCNS-D2; Site\_1: EcoRI; Site\_2: NotI; The poly (A) + RNA was dephosphorylated with bacterial alkaline phosphatase (BAP) and then decapped with tobacco acid pyrophosphatase (TAP). The decapped intact mRNA was ligated with DNA-RNA linker including EcoRI site by treatment of T4 RNA ligase and the first strand cDNA was synthesized from oligo dt-selected mRNA by priming with dt-tailed vector. The dt-tailed vector was adjusted to have about 60nt. The cDNA vector was circularized with E. coli DNA ligase after digestion of EcoRI which site is also included in vector. An RNA strand converted to a DNA strand by Okayama-Berg method. The obtained cDNA vectors were used for transformation of competent cells E. coli Top10F' by electroporation method. The cDNA libraries constructed by this method are full-length enriched cDNA library."

## ORIGIN

Query Match 100.0%; Score 17; DB 4; Length 621;  
Best Local Similarity 100.0%; Pred. No. 5.4e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACGCTTGCCGAG 17  
Db 314 GCCACGCTTGCCGAG 330

RESULT 26  
BE747107  
LOCUS BE747107 630 bp mRNA linear EST 15-SEP-2000  
DEFINITION 601580680F1 NIH\_MGC\_9 Homo sapiens cDNA clone IMAGE:3929706 5', mRNA sequence.  
ACCESSION BE747107  
VERSION BE747107.1 GI:10161099  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 630)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabs-r@mail.nih.gov  
Tissue Procurement: DCTD/DTF  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
Plate: L1CM762 row: 0 column: 19  
High quality sequence stop: 628.  
Location/Qualifiers  
1. .630  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3929706"  
/tissue\_type="adenocarcinoma cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_9"  
/note="Organ: ovary; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

FEATURES  
source

1. .630  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3929706"  
/tissue\_type="adenocarcinoma cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_9"  
/note="Organ: ovary; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

## ORIGIN

Query Match 100.0%; Score 17; DB 7; Length 630;  
Best Local Similarity 100.0%; Pred. No. 5.4e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACGCTTGCCGAG 17  
Db 66 GCCACGCTTGCCGAG 82

## RESULT 27

BM843051  
LOCUS BM843051 637 bp mRNA linear EST 06-MAR-2002  
DEFINITION K-EST0120649 S12SNU216 Homo sapiens cDNA clone S12SNU216-77-H12 5', mRNA sequence.  
ACCESSION BM843051  
VERSION BM843051.1 GI:19199460  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS

1 (bases 1 to 637)  
Oh, J.H., Yang, J.O., Hahn, Y., Kim, M.R., Byun, S.S., Jeon, Y.J., Kim, J.M., Song, K.S., Noh, S.M., Kim, S., Yoo, H.S., Kim, Y.S. and Kim, N.S.  
Transcriptome analysis of human gastric cancer  
Mamm. Genome 16 (12), 942-954 (2005)  
16341674  
Contact: Kim YS  
Genome Research Center  
Korea Research Institute of Bioscience & Biotechnology  
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea  
Tel: +82-42-860-4470  
Fax: +82-42-860-4409  
Email: yongsung@mail.kribb.re.kr  
Plate: 77 row: H column: 12  
High quality sequence stop: 637.  
Location/Qualifiers  
1. .637  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="S12SNU216-77-H12"  
/sex="F"  
/tissue\_type="Lymph node"  
/cell\_type="Epithelial"  
/cell\_line="SNU-216"  
/lab\_host="Top10F"  
/clone\_lib="S12SNU216"  
/note="Organ: Stomach; Vector: pCNS; Site\_1: EcoRI; Site\_2: NotI; The poly (A) + RNA was dephosphorylated with bacterial alkaline phosphatase (BAP) and then decapped with tobacco acid pyrophosphatase (TAP). The decapped intact mRNA was ligated with DNA-RNA linker including EcoRI site by treatment of T4 RNA ligase and the first strand cDNA was synthesized from oligo dt-selected mRNA by priming with dt-tailed vector. The dt-tailed vector was adjusted to have about 60nt. The cDNA vector was circularized with E. coli DNA ligase after digestion of EcoRI which site is also included in vector. An RNA strand converted to a DNA strand by Okayama-Berg method. The obtained cDNA vectors were used for transformation of competent cells E. coli Top10F' by electroporation method. The cDNA libraries constructed by this method are full-length enriched cDNA library."

TITLE  
JOURNAL  
PUBMED  
COMMENT

Transcriptome analysis of human gastric cancer  
Mamm. Genome 16 (12), 942-954 (2005)  
16341674  
Contact: Kim YS  
Genome Research Center  
Korea Research Institute of Bioscience & Biotechnology  
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea  
Tel: +82-42-860-4470  
Fax: +82-42-860-4409  
Email: yongsung@mail.kribb.re.kr  
Plate: 77 row: H column: 12  
High quality sequence stop: 637.  
Location/Qualifiers  
1. .637  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="S12SNU216-77-H12"  
/sex="F"  
/tissue\_type="Lymph node"  
/cell\_type="Epithelial"  
/cell\_line="SNU-216"  
/lab\_host="Top10F"  
/clone\_lib="S12SNU216"  
/note="Organ: Stomach; Vector: pCNS; Site\_1: EcoRI; Site\_2: NotI; The poly (A) + RNA was dephosphorylated with bacterial alkaline phosphatase (BAP) and then decapped with tobacco acid pyrophosphatase (TAP). The decapped intact mRNA was ligated with DNA-RNA linker including EcoRI site by treatment of T4 RNA ligase and the first strand cDNA was synthesized from oligo dt-selected mRNA by priming with dt-tailed vector. The dt-tailed vector was adjusted to have about 60nt. The cDNA vector was circularized with E. coli DNA ligase after digestion of EcoRI which site is also included in vector. An RNA strand converted to a DNA strand by Okayama-Berg method. The obtained cDNA vectors were used for transformation of competent cells E. coli Top10F' by electroporation method. The cDNA libraries constructed by this method are full-length enriched cDNA library."

FEATURES  
source

1. .637  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="S12SNU216-77-H12"  
/sex="F"  
/tissue\_type="Lymph node"  
/cell\_type="Epithelial"  
/cell\_line="SNU-216"  
/lab\_host="Top10F"  
/clone\_lib="S12SNU216"  
/note="Organ: Stomach; Vector: pCNS; Site\_1: EcoRI; Site\_2: NotI; The poly (A) + RNA was dephosphorylated with bacterial alkaline phosphatase (BAP) and then decapped with tobacco acid pyrophosphatase (TAP). The decapped intact mRNA was ligated with DNA-RNA linker including EcoRI site by treatment of T4 RNA ligase and the first strand cDNA was synthesized from oligo dt-selected mRNA by priming with dt-tailed vector. The dt-tailed vector was adjusted to have about 60nt. The cDNA vector was circularized with E. coli DNA ligase after digestion of EcoRI which site is also included in vector. An RNA strand converted to a DNA strand by Okayama-Berg method. The obtained cDNA vectors were used for transformation of competent cells E. coli Top10F' by electroporation method. The cDNA libraries constructed by this method are full-length enriched cDNA library."

## ORIGIN

Query Match 100.0%; Score 17; DB 3; Length 637;  
Best Local Similarity 100.0%; Pred. No. 5.4e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACGCTTGCCGAG 17

```

Db      59 GCCCAGCGTTGGCCGAG 75
|||||
RESULT 28
BF568108
LOCUS    602183908F1 NIH_MGC_42 Homo sapiens cDNA clone IMAGE:4300248 5',
DEFINITION linear mRNA EST 12-DEC-2000
        mRNA sequence.
ACCESSION BF568108
VERSION    602183908F1 GI:11641526
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
          Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
          Hominidae; Homo.
REFERENCE 1 (bases 1 to 699)
AUTHORS  NIH-MGC http://mgc.nci.nih.gov/.
TITLE    National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL  Unpublished (1999)
COMMENT  Contact: Robert Strausberg, Ph.D.
        Email: cgapbs-remail.nih.gov
        Tissue Procurement: ATCC
        cDNA Library Preparation: Ling Hong/Rubin Laboratory
        DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
        DNA Sequencing by: Incyte Genomics, Inc.
        Clone distribution: MGC clone distribution information can be
        found through the I.M.A.G.E. Consortium/LLNL at:
        http://image.llnl.gov
        Plate: LLCM1158 row: 0 column: 01
        High quality sequence stop: 638.
FEATURES
        source
        Location/Qualifiers
          1..699
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:4300248"
            /tissue_type="epithelioid carcinoma cell line"
            /lab_host="DH10B (phage-resistant)"
            /clone_lib="NIH_MGC_42"
            /note="Organ: pancreas; Vector: pOTB7; Site 1: XhoI;
            Site 2: EcoRI; cDNA made by oligo-dT priming.
            Directionally cloned into EcoRI/XhoI sites using the
            following 5' adaptor: GCCAGCG(G). Size-selected >50bp
            for average insert size 1.8kb. Library constructed by Ling
            Hong in the laboratory of Gerald M. Rubin (University of
            California, Berkeley) using ZAP-cDNA synthesis kit
            (Stratagene) and Superscript II RT (Life Technologies).
            Note: this is a NIH_MGC Library."
ORIGIN
Query Match 100.0%; Score 17; DB 7; Length 699;
Best Local Similarity 100.0%; Pred. No. 5.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17
|||||
Db 321 GCCCAGCGTTGGCCGAG 337
|||||
RESULT 29
BG251131
LOCUS    602364932F1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:4473439 5',
DEFINITION linear mRNA EST 13-FEB-2001
        mRNA sequence.
ACCESSION BG251131
VERSION    602364932F1 GI:12760947
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
          Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
          Hominidae; Homo.
REFERENCE 1 (bases 1 to 702)
AUTHORS  NIH-MGC http://mgc.nci.nih.gov/.
TITLE    National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL  Unpublished (1999)
COMMENT  Contact: Robert Strausberg, Ph.D.
        Email: cgapbs-remail.nih.gov
        Tissue Procurement: DCTD/DTP
        cDNA Library Preparation: Rubin Laboratory
        DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
        DNA Sequencing by: Agencourt Bioscience Corporation
        Clone distribution: MGC clone distribution information can be
        found through the I.M.A.G.E. Consortium/LLNL at:
        http://image.llnl.gov
        Plate: LLCM2570 row: 1 column: 09
        High quality sequence stop: 579.
FEATURES
        source
        Location/Qualifiers
          1..702
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:4473439"
            /tissue_type="adenocarcinoma, cell line"
            /lab_host="DH10B (phage-resistant)"
            /clone_lib="NIH_MGC_90"
            /note="Organ: liver; Vector: pCMV-SPORT6; Site 1: NotI;
            Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
            Average insert size 1.7 kb. Library enriched for
            full-length clones and constructed by Life Technologies.
            Note: this is a NIH_MGC Library."
ORIGIN
Query Match 100.0%; Score 17; DB 2; Length 702;
Best Local Similarity 100.0%; Pred. No. 5.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17
|||||
Db 464 GCCCAGCGTTGGCCGAG 480
|||||
RESULT 30
BQ887352
LOCUS    8678064 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6381080
DEFINITION linear mRNA EST 16-AUG-2002
        5', mRNA sequence.
ACCESSION BQ887352
VERSION    8678064 GI:22279366
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
          Hominidae; Homo.
REFERENCE 1 (bases 1 to 851)
AUTHORS  NIH-MGC http://mgc.nci.nih.gov/.
TITLE    National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL  Unpublished (1999)
COMMENT  Contact: Robert Strausberg, Ph.D.
        Email: cgapbs-remail.nih.gov
        Tissue Procurement: DCTD/DTP
        cDNA Library Preparation: Rubin Laboratory
        DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
        DNA Sequencing by: Agencourt Bioscience Corporation
        Clone distribution: MGC clone distribution information can be
        found through the I.M.A.G.E. Consortium/LLNL at:
        http://image.llnl.gov
        Plate: LLCM2570 row: 1 column: 09
        High quality sequence stop: 579.
FEATURES
        source
        Location/Qualifiers
          1..851
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:6381080"
            /tissue_type="adenocarcinoma, cell line"
            /lab_host="DH10B (phage-resistant)"
            /clone_lib="NIH_MGC_40"
            /note="Organ: liver; Vector: pCMV-SPORT6; Site 1: NotI;
            Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
            Average insert size 1.7 kb. Library enriched for
            full-length clones and constructed by Life Technologies.
            Note: this is a NIH_MGC Library."
ORIGIN
Query Match 100.0%; Score 17; DB 2; Length 702;
Best Local Similarity 100.0%; Pred. No. 5.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17
|||||
Db 464 GCCCAGCGTTGGCCGAG 480
|||||
RESULT 30
BQ887352
LOCUS    8678064 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6381080
DEFINITION linear mRNA EST 16-AUG-2002
        5', mRNA sequence.
ACCESSION BQ887352
VERSION    8678064 GI:22279366
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
          Hominidae; Homo.
REFERENCE 1 (bases 1 to 851)
AUTHORS  NIH-MGC http://mgc.nci.nih.gov/.
TITLE    National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL  Unpublished (1999)
COMMENT  Contact: Robert Strausberg, Ph.D.
        Email: cgapbs-remail.nih.gov
        Tissue Procurement: DCTD/DTP
        cDNA Library Preparation: Rubin Laboratory
        DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
        DNA Sequencing by: Agencourt Bioscience Corporation
        Clone distribution: MGC clone distribution information can be
        found through the I.M.A.G.E. Consortium/LLNL at:
        http://image.llnl.gov
        Plate: LLCM2570 row: 1 column: 09
        High quality sequence stop: 579.
FEATURES
        source
        Location/Qualifiers
          1..851
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:6381080"
            /tissue_type="adenocarcinoma, cell line"
            /lab_host="DH10B (phage-resistant)"
            /clone_lib="NIH_MGC_40"
            /note="Organ: liver; Vector: pCMV-SPORT6; Site 1: NotI;
            Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
            Average insert size 1.7 kb. Library enriched for
            full-length clones and constructed by Life Technologies.
            Note: this is a NIH_MGC Library."
ORIGIN
Query Match 100.0%; Score 17; DB 2; Length 702;
Best Local Similarity 100.0%; Pred. No. 5.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17
|||||
Db 464 GCCCAGCGTTGGCCGAG 480
|||||

```

```

/db_xref="taxon:9606"
/clone="IMAGE:6381080"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_40"
/notes="Organ: prostate; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCCAGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

```

## ORIGIN

```

Query Match      100.0%; Score 17; DB 3; Length 851;
Best Local Similarity 100.0%; Pred. No. 5.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17
    |||||
Db 182 GCCCAGCGTTGGCCGAG 198

```

```

RESULT 31
CA454746
LOCUS
DEFINITION
CA454746 865 bp mRNA linear EST 12-NOV-2002
AGENCOURT 10763191 MAPcL Homo sapiens cDNA clone IMAGE:6721245 5',
mRNA sequence.

```

```

ACCESSION
CA454746
VERSION
CA454746.1 GI:24904781
KEYWORDS
EST.
SOURCE
Homo sapiens (human)

```

## ORGANISM

```

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.

```

```

REFERENCE
1 (bases 1 to 865)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)

```

## AUTHORS

```

TITLE
Contact: Robert Strausberg, Ph.D.
Tissue Procurement: Kristi A. Eglund, Ira Pastan
cDNA Library Preparation: Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

```

```

Plate: LHAM14282 row: a column: 21
High quality sequence stop: 686.

```

## FEATURES

## source

```

1..865
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6721245"
/cell_line="ZR-75-1, MCF7, SK-BR-3, MDA-MB-231,
hTERT-HME1, LNCaP"
/lab_host="EMDH10B"
/clone_lib="MAPcL"
/notes="Vector: pCMV-SPORT6; Site_1: EcoRV; Site_2: Not I;
Subtracted with brain, liver, lung, kidney and muscle.
Directionally cloned. Priming method: oligo-dT. Average
insert size: 1800 bp. Library amplification: 26,000 fold.
Kristi A. Eglund, James J. Vincent, Robert Strausberg,
Bungkook Lee & Ira Pastan. Discovery of new breast
cancer genes encoding membrane and secreted proteins.
Manuscript submitted."

```

## ORIGIN

```

Query Match      100.0%; Score 17; DB 4; Length 865;

```

```

Best Local Similarity 100.0%; Pred. No. 5.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17
    |||||
Db 611 GCCCAGCGTTGGCCGAG 627

BM011054
603634744F1 NIH_MGC_47 Homo sapiens cDNA clone IMAGE:5421755 5',
mRNA sequence.

```

## RESULT 32

## BM011054

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

**SOURCE** Homo sapiens (human)  
**ORGANISM** Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Hominidae; Homo.

**REFERENCE** 1 (bases 1 to 899)  
**AUTHORS** NIH-MGC <http://mgs.nci.nih.gov/>.  
**TITLE** National Institutes of Health, Mammalian Gene Collection (MGC)  
**JOURNAL** Unpublished (1999)  
**COMMENT** Contact: Robert Strausberg, Ph.D.  
 Tissue Procurement: DCTD/DTF  
 cDNA Library Preparation: Rubin Laboratory  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLCM2660 row: f column: 02  
 High quality sequence stop: 636.

**FEATURES** Location/Qualifiers  
 source  
 1..899  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6480481"  
 /tissue\_type="carcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_40"  
 /note="Organ: prostate; Vector: pOTB7; Site 1: XhoI;  
 Site 2: EcoRI; cDNA made by oligo-dT priming.  
 Directionally cloned into EcoRI/XhoI sites using the  
 following 5' adaptor: GCCACGAG(G). Library constructed by  
 Ling Hong in the laboratory of Gerald M. Rubin (University  
 of California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies).  
 Note: this is a NIH\_MGC Library."

**ORIGIN**  
 Query Match 100.0%; Score 17; DB 3; Length 899;  
 Best Local Similarity 100.0%; Pred. No. 5.4e+02;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

**Qy** 1 GCCACCGCTTGGCCGAG 17  
 |||||

**Db** 546 GCCACCGCTTGGCCGAG 562  
 |||||

**RESULT 34**  
**LOCUS** BG831564  
**DEFINITION** 6027658139f1 NIH\_MGC\_42 Homo sapiens cDNA clone IMAGE:4907885 5',  
 mRNA sequence.  
**ACCESSION** BG831564  
**VERSION** BG831564.1 GI:14179151  
**KEYWORDS** EST.  
**SOURCE** Homo sapiens (human)  
**ORGANISM** Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Hominidae; Homo.

**REFERENCE** 1 (bases 1 to 901)  
**AUTHORS** NIH-MGC <http://mgs.nci.nih.gov/>.  
**TITLE** National Institutes of Health, Mammalian Gene Collection (MGC)  
**JOURNAL** Unpublished (1999)  
**COMMENT** Contact: Robert Strausberg, Ph.D.  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>  
 Plate: LLCM1813 row: e column: 06  
 High quality sequence stop: 663.

**FEATURES** Location/Qualifiers  
 source  
 1..901  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4907885"  
 /tissue\_type="epithelioid carcinoma cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_42"  
 /note="Organ: pancreas; Vector: pOTB7; Site 1: XhoI;  
 Site 2: EcoRI; cDNA made by oligo-dT priming.  
 Directionally cloned into EcoRI/XhoI sites using the  
 following 5' adaptor: GCCACGAG(G). Size-selected >500bp  
 for average insert size 1.8kb. Library constructed by Ling  
 Hong in the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies).  
 Note: this is a NIH\_MGC Library."

**ORIGIN**  
 Query Match 100.0%; Score 17; DB 2; Length 901;  
 Best Local Similarity 100.0%; Pred. No. 5.4e+02;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

**Qy** 1 GCCACCGCTTGGCCGAG 17  
 |||||

**Db** 322 GCCACCGCTTGGCCGAG 338  
 |||||

**RESULT 35**  
**LOCUS** CF780868  
**DEFINITION** AGENCOURT 15939245 NIH MGC 219 Homo sapiens cDNA clone  
 IMAGE:30523569 5', mRNA sequence.  
**ACCESSION** CF780868  
**VERSION** CF780868.1 GI:37740645  
**KEYWORDS** EST.  
**SOURCE** Homo sapiens (human)  
**ORGANISM** Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Hominidae; Homo.

**REFERENCE** 1 (bases 1 to 916)  
**AUTHORS** NIH-MGC <http://mgs.nci.nih.gov/>.  
**TITLE** National Institutes of Health, Mammalian Gene Collection (MGC)  
**JOURNAL** Unpublished (1999)  
**COMMENT** Contact: Daniela S. Gerhard, Ph.D.  
 Office of Cancer Genomics / NIH  
 National Cancer Institute / NIH  
 Bldg. 31 Rm10A07 Bethesda, MD 20892  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: James Martin, University of Iowa  
 cDNA Library Preparation: M. Bento Soares, University of Iowa  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: NDAM601 row: l column: 10  
 High quality sequence stop: 651.

**FEATURES** Location/Qualifiers  
 source  
 1..916  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:30523569"  
 /tissue\_type="Pooled Chondrosarcoma Tumor cells"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /clone\_lib="NIH\_MGC\_219"  
 /note="Vector: pYX-Asc; Site 1: EcoRI; Site 2: NotI;

Library is oligo-dT primed and directionally cloned. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with Not I and then cloned directionally into pX-Asc vector. Average insert size 0.5-1kb. Adaptors 5'(AATTCGGCAGG)3' and 5'(CTCGTGGCG)3'. 3' Linker sequence - GCGCGCGTGGAGCC T18. Sequencing primers 3'end: T3 promoter primer 5'd (ATTAACCTCACTAAGGA)3'. 5' End: T7 promoter primer 5'd (TAATGACTACTATAGG)3'. Library was constructed in the laboratory of M. Bento Soares. Average insert size 2-3kb. Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 100.0%; Score 17; DB 5; Length 916;  
Best Local Similarity 100.0%; Pred. No. 5.4e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17  
|||||  
Db 667 GCCCAGCGTTGGCCGAG 683

## RESULT 36

BX372160/c

LOCUS BX372160

DEFINITION BX372160 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens cDNA clone CS0DK011YF17 3-PRIME, mRNA sequence.

ACCESSION BX372160

VERSION BX372160.1

KEYWORDS GI:30452082

SOURCE Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 918)

AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.

TITLE Full-length cDNA libraries and normalization

JOURNAL Unpublished (2001)

COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage

2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE

Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 6601.f

For more information about this cluster, see

http://www.genoscope.cns.fr/cdna?s=CS0BAI040ZB07\_CS03797\_1&c=6601.f

## FEATURES

source

Location/Qualifiers

1..918  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clones="CS0DK011YF17"  
/cell\_type="HELA CELLS COT 25-NORMALIZED"  
/cell\_lines="HELA"  
/clone\_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"  
/note="First strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

## ORIGIN

Query Match 100.0%; Score 17; DB 4; Length 918;

Best Local Similarity 100.0%; Pred. No. 5.4e+02;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17  
|||||  
Db 335 GCCCAGCGTTGGCCGAG 319

## RESULT 37

BU543952

LOCUS BU543952

DEFINITION BU543952 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens 5' mRNA sequence.

ACCESSION BU543952

VERSION BU543952.1

KEYWORDS GI:22854435

SOURCE Homo sapiens

ORGANISM

Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 923)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: DCTD/DTF

cDNA Library Preparation: Rubin Laboratory

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLCM2774 row: d column: 08

High quality sequence stop: 641.

## FEATURES

source

Location/Qualifiers

1..923  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6576152"  
/tissue\_type="carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_40"  
/note="Organ: prostate; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 100.0%; Score 17; DB 3; Length 923;

Best Local Similarity 100.0%; Pred. No. 5.4e+02;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17  
|||||  
Db 546 GCCCAGCGTTGGCCGAG 562

## RESULT 38

BX370558

LOCUS BX370558

DEFINITION BX370558 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens cDNA clone CS0DK011YF17 5-PRIME, mRNA sequence.

ACCESSION BX370558

VERSION BX370558.1

KEYWORDS GI:30453927

SOURCE Homo sapiens

ORGANISM

Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
1 (bases 1 to 962)  
Li.W.B., Gruber,C., Jessee,J. and Polayes,D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
end enriched, double-strand cDNA was digested with Not I and cloned  
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
was normalized. Library was constructed by Life Technologies, a  
division of Invitrogen. This sequence belongs to sequence cluster  
6601.f  
For more information about this cluster, see  
http://www.genoscope.cns.fr/cdna?s=CS0BAG0492C03\_CS04627\_lkc=6601.f

# FEATURES

source  
Location/Qualifiers  
1..962  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DK011YF17"  
/cell\_type="HELA CELLS COT 25-NORMALIZED"  
/cell\_line="HELA"  
/clone\_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

Query Match 100.0%; Score 17; DB 4; Length 962;  
Best Local Similarity 100.0%; Pred. No. 5.4e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17  
|||||  
Db 248 GCCCAGCGTTGGCCGAG 264  
|||||

RESULT 39  
BQ067433  
LOCUS  
DEFINITION  
AGENCOURT 6758944 NIH\_MGC\_115 Homo sapiens cDNA clone IMAGE:5754944  
5', mRNA sequence.  
ACCESSION  
BQ067433  
VERSION  
BQ067433.1 GI:19896479  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
1 (bases 1 to 1023)  
NIH-MGC http://mgs.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Life Technologies, Inc.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L1AM12793 row: c column: 09  
High quality sequence stop: 649.  
Location/Qualifiers  
1..1023

# ORIGIN

Query Match 100.0%; Score 17; DB 3; Length 1023;  
Best Local Similarity 100.0%; Pred. No. 5.4e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17  
|||||  
Db 248 GCCCAGCGTTGGCCGAG 264  
|||||

RESULT 40  
B1193620  
LOCUS  
DEFINITION  
B1193620 NIH\_MGC\_42 Homo sapiens cDNA clone IMAGE:5089900 5',  
mRNA sequence.  
ACCESSION  
B1193620  
VERSION  
B1193620.1 GI:14648640  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
1 (bases 1 to 1125)  
NIH-MGC http://mgs.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L1CM1851 row: e column: 05  
High quality sequence stop: 529.  
Location/Qualifiers  
1..1125  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5089900"  
/tissue\_type="epithelioid carcinoma cell line"  
/lab\_host="NIH\_MGC\_42"  
/clone\_lib="NIH\_MGC\_42"  
/note="Organ: pancreas; Vector: pOTB7; Site: 1: XhoI;  
Site 2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCACGAG(G). Size-selected  
for average insert size 1.8kb. Library constructed by Ling  
Hong in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library."

ORIGIN

Query Match 100.0%; Score 17; DB 3; Length 1023;  
Best Local Similarity 100.0%; Pred. No. 5.4e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17  
|||||  
Db 403 GCCCAGCGTTGGCCGAG 419  
|||||

RESULT 40  
B1193620  
LOCUS  
DEFINITION  
B1193620 NIH\_MGC\_42 Homo sapiens cDNA clone IMAGE:5089900 5',  
mRNA sequence.  
ACCESSION  
B1193620  
VERSION  
B1193620.1 GI:14648640  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
1 (bases 1 to 1125)  
NIH-MGC http://mgs.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L1CM1851 row: e column: 05  
High quality sequence stop: 529.  
Location/Qualifiers  
1..1125  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5089900"  
/tissue\_type="epithelioid carcinoma cell line"  
/lab\_host="NIH\_MGC\_42"  
/clone\_lib="NIH\_MGC\_42"  
/note="Organ: pancreas; Vector: pOTB7; Site: 1: XhoI;  
Site 2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCACGAG(G). Size-selected  
for average insert size 1.8kb. Library constructed by Ling  
Hong in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library."

FEATURES  
source

## ORIGIN

Query Match 100.0%; Score 17; DB 2; Length 1125;  
 Best Local Similarity 100.0%; Pred. No. 5.4e+02; Indels 0; Gaps 0;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGGCCGAG 17  
 |||||

Db 294 GCCACGCTTGGCCGAG 310  
 |||||

## RESULT 41

CR597125 1830 bp mRNA linear HTC 21-JUL-2004  
 LOCUS full-length cDNA clone CS0DK011YF17 of HeLa cells Cot 25-normalized  
 DEFINITION of Homo sapiens (human).

ACCESSION CR597125 GI:50477932

VERSION CR597125.1 HTCCNSLT\_CDNA.

KEYWORDS Homo sapiens (human)

SOURCE Homo sapiens

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.

1 (bases 1 to 1830)

Li, W.B., Gruber, C., Jesses, J. and Polayes, D.

Full-length cDNA libraries and normalization

Unpublished

Unpublished

Contact: Feng Liang Email: fliang@lifetech.com URL:

http://fulllength.invitrogen.com/Invitrogen Corporation 1600

Faraday Avenue

Genoscope.

2 (bases 1 to 1830)

Direct Submission

Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage:

BP 191 91006 Evry cedex - FRANCE (E-mail: seqref@genoscope.cns.fr)

- Web: www.genoscope.cns.fr

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime

end enriched, double-strand cDNA was digested with Not I and cloned

into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library

was normalized. Library was constructed by Life Technologies, a

division of Invitrogen.

Location/Qualifiers

1. 1830

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clones="CS0DK011YF17"

/tissue="Hela cells Cot 25-normalized"

/plasmid="pCMVSPORT\_6"

## ORIGIN

Query Match 100.0%; Score 17; DB 6; Length 1830;  
 Best Local Similarity 100.0%; Pred. No. 5.4e+02; Indels 0; Gaps 0;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGGCCGAG 17  
 |||||

Db 1118 GCCACGCTTGGCCGAG 1134  
 |||||

## RESULT 42

AA041289/c 322 bp mRNA linear EST 30-AUG-1996  
 LOCUS zf08d03.e1 Soares fetal heart NBHL19W Homo sapiens cDNA clone  
 DEFINITION IMAGE376325 3' similar to gb:X02162 AF010PROTEIN A-I PRECURSOR  
 (HUMAN); mRNA sequence.

ACCESSION AA041289

VERSION AA041289.1 GI:1517506

KEYWORDS Homo sapiens (human)

SOURCE Homo sapiens

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Euarchontoglires; Primates; Catarrhini;  
 Hominidae; Homo.

1 (bases 1 to 322)

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,

Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,

Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, E.,

Trevaskis, E., Waterston, R., Williamson, A., Wohldmann, P. and

Wilson, R.

The WashU-Merck EST Project

Unpublished (1995)

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -40M13 fwd. from Amersham

High quality sequence stop: 288.

## FEATURES

source

1. 322

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="GDB:1284581"

/db\_xref="taxon:9606"

/clone="IMAGE:376325"

/sex="unknown"

/dev\_stage="19 weeks"

/lab\_host="DH10B (ampicillin resistant)"

/clone\_lib="Soares fetal heart NBHL19W"

/notes="Organ: heart; Vector: pT7T3D (Pharmacia) with a

modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st

strand cDNA was primed with a Not I - oligo(dT) primer [5'

TGTTACCAATCTGAGTGGGCGCGCATCTTTT TTTT TTTT 3']

double-stranded cDNA was size selected, ligated to Eco RI

adapters (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of a modified pT7T3 vector

(Pharmacia). Library went through one round of

normalization to a Cot = 5. Library constructed by

M.Fatima Bonaldo. This library was constructed from the

same fetus as the fetal lung library, Soares fetal lung

NBHL19W."

## ORIGIN

Query Match 90.6%; Score 15.4; DB 1; Length 322;  
 Best Local Similarity 94.1%; Pred. No. 3.6e+03;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGGCCGAG 17  
 |||||

Db 270 GCCACGCTTGGCCGCG 254  
 |||||

## RESULT 43

BY153767/c

LOCUS

DEFINITION

BY153767 RIKEN full-length enriched, adult pancreas islet cells Mus

musculus cDNA clone C820014J12 5', mRNA sequence.

ACCESSION BY153767

VERSION BY153767.1 GI:26290313

KEYWORDS EST.

SOURCE

Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 363)

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,

Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I.,

Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,

Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C.,

Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,



Batalov, S., Beisel, K.W., Blake, J.A., Bratt, D., Brusic, V., Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawaji, H., Kawaasa, Y., Kedzierski, R.M., King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maltate, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Perte, G., Peole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramchandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavalan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Aizawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
12466851

Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.jp, URL: <http://genome.gsc.riken.jp/>

Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in RIKEN Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Cells were provided by Hiroo Iwata (Department of Reproductive Materials Field of Tissue Engineering Institute for Frontier Medical Sciences, Sakyo-ku, Kyoto, 606-8507, Japan) whose assistance we gratefully acknowledge.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

Location/Qualifiers  
1. .363  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="CB20014J12"  
/tissue\_type="pancreas"  
/cell\_type="islet cells"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/clone\_lib="RIKEN full-length enriched, adult pancreas

islet cells"  
/note="Site 1: Sali; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'  
GAGAGAGAGCGCGCCCACTCGAGTTTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5'  
GAGAGAGAGATTCTCGAGTTTAATAATTAATCCCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cells were provided by Tomohiro Kono, Department of Animal Science, Tokyo University of Agriculture, Atsugi City, Kanagawa Prefecture, Japan, whose assistance we gratefully acknowledge."

## ORIGIN

Query Match 90.6%; Score 15.4; DB 4; Length 363;  
Best Local Similarity 94.1%; Pred. No. 3.6e+03;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTGGCCGAG 17  
|||||  
Db 206 GCCCGCGCTGGCCGAG 190  
|||||

RESULT 44  
BY774337/c

LOCUS BY774337 365 bp mRNA linear EST 23-MAR-2004  
DEFINITION BY774337 RIKEN full-length enriched, 17.5 days embryo whole body  
Mus musculus cDNA clone L930083C03 5', mRNA sequence.  
BY774337  
ACCESSION BY774337.1 GI:39700975  
VERSION EST.  
KEYWORDS Mus musculus (house mouse)  
SOURCE

ORGANISM  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Murinae; Mus.

## REFERENCE

AUTHORS  
Carninci, P., Waki, K., Shiraki, T., Konno, H., Shibata, K., Itoh, M., Aizawa, K., Arakawa, T., Ishii, Y., Sasaki, D., Bono, H., Kondo, S., Sugahara, Y., Saito, R., Osato, N., Fukuda, S., Sato, K., Watahiki, A., Hirozane-Kishikawa, T., Nakamura, M., Shibata, Y., Yasunishi, A., Kikuchi, N., Yoshiki, A., Kusakabe, M., Gustincich, S., Beisel, K., Pavan, W., Aidinis, V., Nakagawara, A., Held, W.A., Iwata, H., Kono, T., Nakachi, H., Lyons, P., Wells, C., Hume, D.A., Fagioli, M., Hensch, T.K., Brinkmeier, M., Camper, S., Hirota, J., Mombaerts, P., Muramatsu, M., Okazaki, Y., Kawai, J. and Hayashizaki, Y.

## TITLE

Targeting a complex transcriptome: the construction of the mouse full-length cDNA encyclopedia

## JOURNAL

Genome Res. 13 (6B), 1273-1289 (2003)

## PUBMED

12819125

## COMMENT

Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.jp, URL: <http://genome.gsc.riken.jp/>  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Please visit our web site (<http://genome.gsc.riken.jp/>) for further details.

## FEATURES

source

```

FEATURES
  source
    Location/Qualifiers
      1. .365
        /organism="Mus musculus"
        /mol_type="mRNA"
        /strain="C57BL/6J"
        /db_xref="taxon:10090"
        /clone="U930083C03"
        /tissue_type="whole body"
        /dev_stage="17.5 days embryo"
        /clone_lib="RIKEN full-length enriched, 17.5 days embryo
        whole body"

ORIGIN
  Query Match      90.6%; Score 15.4; DB 4; Length 365;
  Best Local Similarity 94.1%; Pred. No. 3.6e+03;
  Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

  QY 1 GCCACGCTTGCGCGAG 17
      ||||| ||||| |||||
  Db 164 GCCCGCGCTTGCGCGAG 148

RESULT 45
BE818831/c
LOCUS
DEFINITION
  B818831 365 bp mRNA linear EST 21-SEP-2000
ACCESSION
  B818831.1 GI:10251065
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
  Hominoidea; Homo.
REFERENCE
  1 (bases 1 to 365)
  Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
  Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
  Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
  Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
  O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
  Simpson,A.J.
  Shotgun sequencing of the human transcriptome with ORF expressed
  sequence tags
  Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
10737800
JOURNAL
  PubMed
COMMENT
  Contact: Simpson A.J.G.
  Laboratory of Cancer Genetics
  Ludwig Institute for Cancer Research
  Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
  Brazil
  Tel: +55-11-2704922
  Fax: +55-11-2707001
  Email: asimpson@ludwig.org.br
  This sequence was derived from the FAPESP/LICR Human Cancer Genome
  Project. This entry can be seen in the following URL
  (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2=CM2-EN0302-050
  700-256-a12&t3=2000-07-05&t4=1)
  Seq primer: puc 18 forward
  High quality sequence start: 14
  High quality sequence stop: 365.
  Location/Qualifiers
    1. .365
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /dev_stage="adult"
      /clone_lib="EN0302"
      /notes="Organ: breast normal; Vector: puc18; Site:1: Smal;
      Site 2: Smal; A mini-library was made by cloning products
      derived from ORESTES PCR (U.S. Letters Patent application
      No. 196, 716 - Ludwig Institute for Cancer Research)
      profiles into the pUC 18 vector. Reverse transcription of
      tissue mRNA and cDNA amplification were performed under

FEATURES
  source
    Location/Qualifiers
      1. .365
        /organism="Mus musculus"
        /mol_type="mRNA"
        /strain="C57BL/6J"
        /db_xref="taxon:10090"
        /clone="U930083C03"
        /tissue_type="whole body"
        /dev_stage="17.5 days embryo"
        /clone_lib="RIKEN full-length enriched, 17.5 days embryo
        whole body"

ORIGIN
  Query Match      90.6%; Score 15.4; DB 7; Length 365;
  Best Local Similarity 94.1%; Pred. No. 3.6e+03;
  Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

  QY 1 GCCACGCTTGCGCGAG 17
      ||||| ||||| |||||
  Db 267 GCCACGCTTGCGCGAG 251

RESULT 46
BY674396
LOCUS
DEFINITION
  BY674396 RIKEN full-length enriched, 14.5 days embryo df/df
  Rathke's pouches Mus musculus cDNA clone K820040J08 3', mRNA
  sequence.
  BY674396
  BY674396.1 GI:27058918
  EST.
  Mus musculus (house mouse)
  Mus musculus
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
  Sciurognathi; Muroidea; Muridae; Murinae; Mus.
  1 (bases 1 to 367)
  Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
  Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaoka,I.,
  Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
  Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
  Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
  Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
  Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
  Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
  Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
  Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
  Kawai,H., Kawasawa,Y., Kedzierski,R., King,B.L., Konggayya,A.,
  Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
  Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
  Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
  Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
  Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
  Sandelin,A., Schneider,C., Sempie,C.A., Setou,M., Shimada,K.,
  Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
  Verardo,R., Wagner,L., Wahlstedt,C., Wang,Y., Watanabe,Y.,
  Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
  Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
  Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
  Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
  Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
  Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
  Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
  Rogers,J., Birney,E. and Hayashizaki,Y.
  Analysis of the mouse transcriptome based on functional annotation
  of 60,770 full-length cDNAs
  Nature 420, 563-573 (2002)
12466851
JOURNAL
  PubMed
COMMENT
  Contact: Yoshihide Hayashizaki
  Laboratory for Genome Exploration Research Group, RIKEN Genomic
  Sciences Center (GSC), Yokohama Institute
  The Institute of Physical and Chemical Research (RIKEN)
  1-7-22 Suehiro-cho, Teurumi-ku, Yokohama, Kanagawa 230-0045, Japan
  Tel: 81-45-503-9222
  Fax: 81-45-503-9216
  Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
  Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,
  Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H.,
  Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R.,
  Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K.,
  Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and
  Hayashizaki,Y. Direct Submission
  Computational Analysis of Full-Length Mouse cDNAs Compared with

```



```

AW206921/c
LOCUS
DEFINITION
  UI-H-B11-afg-d-01-o-UI.s1 NCI CGAP_Sub3 Homo sapiens cDNA clone
  IMAGE:2722968 3', mRNA sequence.
ACCESSION
  AW206921
VERSION
  AW206921.1 GI:6506417
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
ORGANISM
  Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
  Homiidae; Homo.
REFERENCE
  1 (bases 1 to 410)
  NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
  Tumor Gene Index
  Unpublished (1997)
  Contact: Robert Strausberg, Ph.D.
  Email: cgaps-rc@mail.nih.gov
  The sequence contained an oligo-dT track that was present in the
  oligonucleotide that was used to prime the synthesis of first
  strand cDNA and therefore this may represent a bonafide poly A
  tail. cDNA library Preparation: M.B. Soares Lab Clone distribution:
  NCI-CGAP clone distribution information can be found through the
  I.M.A.G.E. Consortium/LLNL at:
  www-bio.llnl.gov/bbrp/image/image.html
  Seq primer: M13 Forward
  POLYA=Yes.
FEATURES
  source
    Location/Qualifiers
      1..410
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /clone="IMAGE:2722968"
        /lab_host="DH10B (Life Technologies)"
        /clone_lib="NCI CGAP Sub3"
        /notes="vector: pVT73D-PacI; Site 1: Not I; Site 2: Eco RI;
        The NCI CGAP Sub3 library is a subtracted library derived
        from the NCI CGAP Sub1 library, which is a subtracted
        library derived from BI. BI constitutes a mixture of 21
        normalized or subtracted NCI CGAP libraries:
        NCI CGAP Co4, NCI CGAP Pr22, NCI CGAP Pr28, NCI CGAP Co10,
        NCI CGAP Co16, NCI CGAP Kid5, NCI CGAP Kid12,
        NCI CGAP Kid3, NCI CGAP Kid11, NCI CGAP Lym2,
        NCI CGAP Br2, NCI CGAP Co8, NCI CGAP CLL1, NCI CGAP Lei2,
        NCI CGAP Brn23, NCI CGAP Lu5, NCI CGAP Lu24,
        NCI CGAP Lu19, NCI CGAP GC4, NCI CGAP GC6,
        NCI CGAP Brn35. These 21 libraries were pooled and a
        single-stranded DNA preparation of the resulting mixture
        was used as a tracer in a subtractive hybridization with
        a driver whose composition is detailed below:
        NCI CGAP Kid3 pool 1 LLAM 3334-3337, 3682-3683,
        3798-3803 (IMAGE CloneIds 132376-132391),
        1456008-1456775, 1500552-1502855; NCI CGAP Kid5 pool 1
        LLAM 3338-3342, 3722-3725, 3776-3778 (IMAGE CloneIds
        1323912-1325831, 1471368-1472903, 1492104-1493255);
        NCI CGAP Lu5 pool 1 LLAM 3575-3582, 3851-3854 (IMAGE
        CloneIds 1414920-1417991, 1520904-1522439); NCI CGAP GC4
        pool 1 LLAM 3164-3167, 3716-3720, 3733-3735 (IMAGE
        CloneIds 1257096-1258631, 1469064-1470983,
        1475592-1476743); NCI CGAP Pr22 pool 1 LLAM 2457-2459,
        2758-2759, 3062-3068 (IMAGE CloneIds 985608-986759,
        1101192-1101959, 1217928-1220615); NCI CGAP Co10 pool 1
        LLAM 2644-2653, 2871-2872 (IMAGE CloneIds 1057416-1061255,
        1144584-1145351). Subtraction was performed as previously
        described [Bonaldi, Lennon & Soares (1996): Normalization
        and Subtraction: Two Approaches to Facilitate Gene
        Discovery. Genome Research 6, 791-806.
        TAG_TISSUE=colon
        TAG_LIB=NCI CGAP_Co10
        TAG_SEQ=AAAGG"

```

```

Query Match      90.6%; Score 15.4; DB 7; Length 410;
Best Local Similarity 94.1%; Pred. No. 3.6e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCGCTTGGCCGAG 17
   |||||
Db 359 GCCCAGCGCTTGGCCGG 343

RESULT 49
BE818807/c
LOCUS
DEFINITION
  CM2-BN0302-040700-253-d02 BN0302 Homo sapiens cDNA, mRNA sequence.
ACCESSION
  BE818807
VERSION
  BE818807.1 GI:10251041
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
  Homiidae; Homo.
REFERENCE
  1 (bases 1 to 449)
  Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
  Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
  Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
  Brunstein,A., deOliveira,P.S., Bucher,F., Jongeneel,C.V.,
  O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
  Simpson,A.J.
  Shotgun sequencing of the human transcriptome with ORF expressed
  sequence tags
  Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  10737800
  Contact: Simpson A.J.G.
  Laboratory of Cancer Genetics
  Ludwig Institute for Cancer Research
  Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
  Brazil
  Tel: +55-11-2704922
  Fax: +55-11-2707001
  Email: asimpson@ludwig.org.br
  This sequence was derived from the FAPESP/LICR Human Cancer Genome
  Project. This entry can be seen in the following URL
  (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2-cm2-bn0302-040)
  700-253-d02&t3=2000-07-04&t4=1)
  Seq primer: puc 18 forward
  High quality sequence start: 24
  High quality sequence stop: 306.
FEATURES
  source
    Location/Qualifiers
      1..449
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /dev_stage="Adult"
        /clone_lib="BN0302"
        /note="Organ: breast normal; Vector: puc18; Site 1: SmaI;
        Site 2: SmaI; A mini-library was made by cloning products
        derived from ORESTIS PCR (U.S. Letters Patent application
        No. 136,716 - Ludwig Institute for Cancer Research)
        profiles into the pUC 18 vector. Reverse transcription of
        tissue mRNA and cDNA amplification were performed under
        low stringency conditions."
ORIGIN
Query Match      90.6%; Score 15.4; DB 7; Length 449;
Best Local Similarity 94.1%; Pred. No. 3.6e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCGCTTGGCCGAG 17
   |||||
Db 240 GCCCAGCGCTTGGCCGG 224

RESULT 50

```

```

DU607602
LOCUS      DU607602              451 bp    DNA        linear    GSS 12-OCT-2005
DEFINITION OO_Ba0106010.r OO_Ba Oryza officinalis genomic clone
ACCESSION  DU607602
VERSION    DU607602.1   GI:77571132
KEYWORDS   GSS.
SOURCE     Oryza officinalis
ORGANISM   Oryza officinalis
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP
            clade; Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE  1 (bases 1 to 451)
AUTHORS   Kim,H., Collura,K., Wisotski,M., Byrne,M., Stum,D., Smart,D.,
            Rao,K., Luo,M., Jetty,R., Kudrna,D., Muller,C., Soderlund,C. and
            Wing,R.
TITLE     OMAP (Oryza Map Alignment Project)- Arizona Genomics Institute
JOURNAL   Unpublished (2005)
COMMENT   Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: rwing@genome.arizona.edu
PCR Primers
FORWARD: TAA TAC GAC TCA CTA TAG GG
BACKWARD: CAC TCA TTA GGC ACC CCA
Plate: 0106 row: O column: 10
Seq primer: CAC TCA TTA GGC ACC CCA
Class: BAC ends.
FEATURES             source
    source            Location/Qualifiers
        1..451
            /organism="Oryza officinalis"
            /mol_type="genomic DNA"
            /db_xref="taxon:4535"
            /clone="OO_Ba0106010"
            /tissue_type="leaves"
            /lab_host="DH10B"
            /clone_lib="OO_Ba"
            /note="vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"

ORIGIN
Query Match      90.6%; Score 15.4; DB 14; Length 451;
Best Local Similarity 94.1%; Pred. No. 3.6e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1  GCCCAGCGTTGGCCGAG 17
        |||||
Db      290 GCCCAGCGTTGGCCGTG 306

RESULT 51
BO965719
LOCUS      BO965719              494 bp    mRNA        linear    EST 21-AUG-2002
DEFINITION QB22M12.VG.ab1 QH ABCDI sunflower RHA801 Helianthus annuus cDNA
            clone QB22M12, mRNA sequence.
ACCESSION  BO965719
VERSION    BO965719.1   GI:22382824
KEYWORDS   EST.
SOURCE     Helianthus annuus (common sunflower)
ORGANISM   Helianthus annuus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; campanulids; Asterales; Asteraceae; Asteroideae;
            Heliantheae; Helianthus.
REFERENCE  1 (bases 1 to 494)
AUTHORS   Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
            Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J.,
            Ellison,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y.,
            Lai,Z., Church,S., Jackson,L. and Bradford,K.
            Lettuce and Sunflower ESTs from the Compositae Genome Project
            http://compgenomics.ucdavis.edu/

```

```

Unpublished (2002)
Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Asmundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@ucdavis.org [michelmore@vegmail.ucdavis.edu]
belongs to contig QH_CA_Contig980, see http://cgdb.ucdavis.edu/
for details.
Plate: QB22 row: M column: 12.
FEATURES             source
    source            Location/Qualifiers
        1..494
            /organism="Helianthus annuus"
            /mol_type="mRNA"
            /cultivar="RHA801"
            /db_xref="taxon:4232"
            /clone="QB22M12"
            /lab_host="E.Coli"
            /clone_lib="QH ABCDI sunflower RHA801"
            /note="vector: pBRCNASFIAB; The library was constructed
            from 11 different sources of RNA from a single genotype.
            Separate cDNAs were generated using primers that
            incorporated unique 5' and 3' tags to distinguish each
            source of RNA. cDNAs were then pooled, size-fractionated,
            directionally cloned into a custom medium-copy vector and
            transformations made with four size classes to minimize
            size bias. Details of each source of RNA and library
            construction can be obtained at http://cgdb.ucdavis.edu/
            TAG_SEQ=Not found"

ORIGIN
Query Match      90.6%; Score 15.4; DB 3; Length 494;
Best Local Similarity 94.1%; Pred. No. 3.6e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1  GCCCAGCGTTGGCCGAG 17
        |||||
Db      465 GCCCAGCGTTGGCCGAG 481

RESULT 52
DA945234
LOCUS      DA945234              566 bp    mRNA        linear    EST 15-NOV-2005
DEFINITION DA945234 SPLEN2 Homo sapiens cDNA clone SPLEN2020625 5', mRNA
            sequence.
ACCESSION  DA945234
VERSION    DA945234.1   GI:82439583
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominoidea; Homo.
REFERENCE  1 (bases 1 to 566)
AUTHORS   Kimura,K., Wakanatsu,A., Suzuki,Y., Ota,T., Nishikawa,T.,
            Yamashita,R., Yamamoto,J., Sekine,M., Tsuritani,K., Wakaguri,H.,
            Ishii,S., Sugiyama,T., Saito,K., Isono,Y., Irie,R., Kushida,N.,
            Yoneyama,T., Otsuka,R., Kanda,K., Yokoi,T., Kondo,H., Wagatsuma,M.,
            Murakawa,K., Ishida,S., Ishibashi,T., Takahashi-Fujii,A.,
            Tanase,T., Nagai,K., Kikuchi,H., Nakai,K., Isogai,T. and Sugano,S.
            Diversification of Transcriptional Modulation: Large-scale
            Identification and Characterization of Putative Alternative
            Promoters of Human Genes
            Genome Res. 16 (1), 55-65 (2006)
            16344560
CONTACT: Takao Isogai
            FLJ Project (HRI Team)
            Helix Research Institute
            2-6-7 Kazusa-Kamatari, Kisarazu, Chiba, 292-0818, Japan
            Tel: 81-438-52-3975
            Fax: 81-438-52-3986
            Email: flj-cdna@nifty.com

```

NEDO human cDNA project (New Energy and Industrial Technology Developmental Organization, Japan); cDNA library construction: Helix Research Institute (HRI); 5'-end one pass sequencing: HRI, Research Association for Biotechnology (RAB) and Biotechnology Center, National Institute of Technology and Evaluation; 3'-end one pass sequencing: RAB. The reason of remaining 5'-end 'N': This sequence was generated from 5'-end region of cDNA insert including about 20 base of the sequence before 'N'. This cDNA was obtained from oligo-capped cDNA library by which method 5'-end of mRNA was efficiently cloned. Therefore remaining 5'-end 'N' is important for a detection of transcription start site of this mRNA.

#### FEATURES

source  
 1. .566  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="SPLEN2020625"  
 /tissue\_type="spleen"  
 /clone\_lib="SPLEN2"  
 /notes="Vector: pME18SFL3"

#### ORIGIN

Query Match 90.6%; Score 15.4; DB 9; Length 566;  
 Best Local Similarity 94.1%; Pred. No. 3.6e+03;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCCAGCCTTGGCCGAG 17

Db 482 GCCCAGCCTTGGCCGAG 498

#### RESULT 53

DA932540  
 LOCUS DA932540 SPLEN2 Homo sapiens cDNA clone SPLEN2002176 5', mRNA  
 DEFINITION sequence.

VERSION DA932540

KEYWORDS DA932540.1 GI:82096064

SOURCE EST.

ORGANISM Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 570)

AUTHORS Kimura,K., Wakamatsu,A., Suzuki,Y., Ota,T., Nishikawa,T., Yamashita,R., Yamamoto,J., Sekine,M., Tsuritani,K., Wakaguri,H., Ishii,S., Sugiyama,T., Saito,K., Isono,Y., Irie,R., Kushida,N., Yoneyama,T., Otsuka,R., Kanda,K., Yokoi,T., Kondo,H., Wagatsuma,M., Murakawa,K., Iehida,S., Ishibashi,T., Takahashi-Fujii,A., Tanase,T., Nagai,K., Kikuchi,H., Nakai,K., Isogai,T. and Sugano,S.  
 TITLE Diversification of Transcriptional Modulation: Large-scale Identification and Characterization of Putative Alternative Promoters of Human Genes

JOURNAL Genome Res. 16 (1), 55-65 (2006)

PUBMED 16344560

COMMENT Contact: Takao Isogai

FLJ Project (HRI Team)

Helix Research Institute

2-6-7 Kazusa-Kamatari, Kisarazu, Chiba, 292-0818, Japan

Tel: 81-438-52-3975

Fax: 81-438-52-3986

Email: flj-cdna@nifty.com

NEDO human cDNA project (New Energy and Industrial Technology Developmental Organization, Japan); cDNA library construction: Helix Research Institute (HRI); 5'-end one pass sequencing: HRI, Research Association for Biotechnology (RAB) and Biotechnology Center, National Institute of Technology and Evaluation; 3'-end one pass sequencing: RAB.

#### FEATURES

source  
 1. .570  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"

/db\_xref="taxon:9606"  
 /clone="SPLEN2002176"  
 /tissue\_type="spleen"  
 /clone\_lib="SPLEN2"  
 /note="Vector: pME18SFL3"

#### ORIGIN

Query Match 90.6%; Score 15.4; DB 9; Length 570;  
 Best Local Similarity 94.1%; Pred. No. 3.6e+03;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCCAGCCTTGGCCGAG 17

Db 459 GCCCAGCCTTGGCCGAG 475

#### RESULT 54

CK783403/c

LOCUS CK783403

DEFINITION CK783403.1 GI:42749081

VERSION EST.

KEYWORDS Alexandrium tamarense

SOURCE Alexandrium tamarense

ORGANISM Eukaryota; Alveolata; Dinophyceae; Gonyaulacales; Gonyaulacaceae; Alexandrium.

REFERENCE 1 (bases 1 to 577)

AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.

TITLE Normalization and subtraction: two approaches to facilitate gene discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)

PUBMED 8889548

COMMENT Contact: Soares, MB

Coordinated Laboratory for Computational Genomics

University of Iowa

375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: bento-soares@uiowa.edu

Tissue Procurement: Provasoli-Guillard National Center for Culture

of Marine Phytoplankton (CCMP)

cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa

cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Distribution information can be found at

http://genome.uiowa.edu/distribution/dinoflagellate.html

Seq primer: M13 FORWARD

POLYA=Yes.

Location/Qualifiers

1. .577

/organism="Alexandrium tamarense"

/mol\_type="mRNA"

/strain="CCMP 1598"

/db\_xref="taxon:2926"

/clone="UI-D-GC1-aau-g-20-0-UI"

/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"

/clone\_lib="UI-D-GC1"

/note="Vector: pT7T3-Pac (Pharmacia) with a modified

polylinker; Site\_1: EcoR I; Site\_2: Not I; UI-D-GC1 is a

normalized library derived from UI-D-GC0. The library was

constructed according to Bonaldo, Lennon and Soares,

Genome Research, 6:791-806, 1996. First strand cDNA

synthesis was primed with an oligo-dT primer containing a

Not I site. Double stranded cDNA was ligated to an EcoR I

adaptor, digested with Not I, and cloned directionally

into pT7T3-Pac vector. The oligonucleotide used to prime

the synthesis of first-strand cDNA contains a library tag

sequence that is located between the Not I site and the

(dT)18 tail. The sequence tag for this library is

TACCTCGAGA. Tissue was obtained from the

Provasoli-Guillard National Center for Culture of Marine



Tanase,T., Nagai,K., Kikuchi,H., Nakai,K., Isogai,T. and Sugano,S.  
 Diversification of Transcriptional Modulation: Large-scale  
 Identification and Characterization of Putative Alternative  
 Promoters of Human Genes  
 Genome Res. 16 (1), 55-65 (2006)  
 16344560  
 Contact: Takao Isogai  
 FLJ Project (HRI Team)  
 Helix Research Institute  
 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba, 292-0818, Japan  
 Tel: 81-438-52-3975  
 Fax: 81-438-52-3986  
 Email: flj-cdna@nifty.com  
 NEDO human cDNA project (New Energy and Industrial Technology  
 Developmental Organization, Japan); cDNA library construction:  
 Helix Research Institute (HRI); 5'-end one pass sequencing: HRI,  
 Research Association for Biotechnology (RAB) and Biotechnology  
 Center, National Institute of Technology and Evaluation; 3'-end one  
 pass sequencing: RAB.

FEATURES  
 source  
 Location/Qualifiers  
 1. 593  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="SPLEN2015028"  
 /tissue\_type="spleen"  
 /clone\_lib="SPLEN2"  
 /note="Vector: pME18SFL3"

ORIGIN  
 Query Match 90.6%; Score 15.4; DB 9; Length 593;  
 Best Local Similarity 94.1%; Pred. No. 3.6e+03;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 1 GCCACGCTTGCGCGAG 17  
 ||||| ||||| ||||| |||||  
 Db 454 GCCACACCTTGCGCGAG 470

RESULT 58  
 CF751855/c  
 LOCUS  
 DEFINITION UI-D-GC0-aaa-b-23-0-UI.s1 600 bp mRNA linear EST 10-OCT-2003  
 UI-D-GC0-aaa-b-23-0-UI 3', mRNA sequence.  
 CF751855  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 EST.  
 Alexandrium tanarense  
 Alexandrium tanarense  
 Eukaryota; Alveolata; Dinophyceae; Gonyaulacales; Gonyaulacaceae;  
 Alexandrium.  
 1 (bases 1 to 600)  
 REFERENCE  
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.  
 TITLE Normalization and subtraction: two approaches to facilitate gene  
 discovery  
 Genome Res. 6 (9), 791-806 (1996)  
 8889548  
 CONTACT: Soares, MB  
 Coordinated Laboratory for Computational Genomics  
 University of Iowa  
 375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA  
 Tel: 319 335 8250  
 Fax: 319 335 9565  
 Email: bento-soares@uiowa.edu  
 Tissue Procurement: Provasoli-Guillard National Center for Culture  
 of Marine Phytoplankton (CCMP)  
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Distribution information can be found at  
<http://genome.uiowa.edu/distribution/dinoflagellate.html>  
 Seq primer: M13 FORWARD  
 POLYA=Yes.

Tanase,T., Nagai,K., Kikuchi,H., Nakai,K., Isogai,T. and Sugano,S.  
 Diversification of Transcriptional Modulation: Large-scale  
 Identification and Characterization of Putative Alternative  
 Promoters of Human Genes  
 Genome Res. 16 (1), 55-65 (2006)  
 16344560  
 Contact: Takao Isogai  
 FLJ Project (HRI Team)  
 Helix Research Institute  
 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba, 292-0818, Japan  
 Tel: 81-438-52-3975  
 Fax: 81-438-52-3986  
 Email: flj-cdna@nifty.com  
 NEDO human cDNA project (New Energy and Industrial Technology  
 Developmental Organization, Japan); cDNA library construction:  
 Helix Research Institute (HRI); 5'-end one pass sequencing: HRI,  
 Research Association for Biotechnology (RAB) and Biotechnology  
 Center, National Institute of Technology and Evaluation; 3'-end one  
 pass sequencing: RAB.

FEATURES  
 source  
 Location/Qualifiers  
 1. 600  
 /organism="Alexandrium tamarense"  
 /mol\_type="mRNA"  
 /strain="CCMP 1598"  
 /db\_xref="taxon:2926"  
 /clone="UI-D-GC0-aaa-b-23-0-UI"  
 /lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
 /clone\_lib="UI-D-GC0"  
 /note="Vector: pT73-Pac (Pharmacia) with a modified  
 polylinker; Site 1: EcoR I; Site 2: Not I; The library was  
 constructed according to Bonaldo, Lennon and Soares,  
 Genome Research, 6:791-806, 1996. First strand cDNA  
 synthesis was primed with an oligo-dT primer containing a  
 Not I site. Double stranded cDNA was ligated to an EcoR I  
 adaptor, digested with Not I, and cloned directionally  
 into pT73-Pac vector. The oligonucleotide used to prime  
 the synthesis of first-strand cDNA contains a library tag  
 sequence that is located between the Not I site and the  
 (dT)18 tail. The sequence tag for this library is  
 TACCTCGAGA. Tissue was obtained from the  
 Provasoli-Guillard National Center for Culture of Marine  
 Phytoplankton (CCMP).  
 TAG\_TISSUE=Alexandrium tanarense  
 TAG\_LIB=UI-D-GC0  
 TAG\_SEQ=TACCTCGAGA"

ORIGIN  
 Query Match 90.6%; Score 15.4; DB 5; Length 600;  
 Best Local Similarity 94.1%; Pred. No. 3.6e+03;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 1 GCCACGCTTGCGCGAG 17  
 ||||| ||||| ||||| |||||  
 Db 74 GCTCAGCGTTGGCGGAG 58

RESULT 59  
 CO196877  
 LOCUS  
 DEFINITION GEO1.2\_F06.bl\_A029 Root gravitropism April 2003 test Pinus taeda  
 cDNA clone GEO1.2\_F06\_A029 3', mRNA sequence.  
 CO196877  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 EST.  
 Pinus taeda (loblolly pine)  
 Pinus taeda  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.  
 1 (bases 1 to 601)  
 REFERENCE  
 AUTHORS Pratt,L., Cordonnier-Pratt,M.-M., Lorenz,W.W., Zimmermann,C. and  
 Dean,J.F.D.  
 TITLE An EST database from gravitropically stimulated loblolly pine  
 (Pinus taeda) roots  
 Unpublished (2004)  
 JOURNAL  
 COMMENT  
 Other\_ESTs: GEO1.2\_F06.g1\_A029  
 Contact: Cordonnier-Pratt MM  
 Laboratory for Genomics and Bioinformatics  
 The University of Georgia, Department of Plant Biology  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 583 0210  
 Email: mmpratt@uga.edu  
 RNA prepared and library constructed by W. Walter Lorenz (School of  
 Forest Resources, University of Georgia); plant material prepared  
 by Craig Zimmermann (School of Forest Resources, University of  
 Georgia) using rooted cuttings provided by the Forest Biology  
 Research Cooperative (FBRC) and the CCLONES project a the  
 University of Florida; sequencing done in the Laboratory for  
 Genomics and Bioinformatics, University of Georgia. Sequence ends  
 have been trimmed to exclude vector and regions below Phred quality  
 16. Three-prime sequences are presented as their reverse complement  
 and have been trimmed to exclude polyA.



Seq primer: ML3-21 (TGTAACAGCAGCGCCAGT)

POLYA=No.

## FEATURES

source

Location/Qualifiers

1..601

/organism="Pinus taeda"

/mol\_type="mRNA"

/strain="3 CCLONES"

/db\_xref="taxon:3352"

/clone="GE01.2 F06 A029"

/lab\_host="DH10B-TI phage-resistant E. coli"

/clone\_lib="Root gravitropism April 2003 test"

/note="Vector: pSL1180; Site 1: EcoRI; Site 2: XhoI; The library was prepared from polyA+ RNA from the roots of 1-year-old loblolly pine (Pinus taeda) cuttings that were rooted and then planted in washed sand. The rooted cuttings were maintained for 27 days (April 2003 harvest) under ambient conditions in a local greenhouse. They were kept on a weekly regimen of 0.5x nutrient-complete Hoagland's solution and supplemented with additional water sufficient to maintain a 15% soil moisture content. Twenty-four hours (24 h) prior to harvesting roots for mRNA preparation, the potted trees were tipped 90 degree to lay the pots on their sides to induce a gravitropic response. Double-stranded cDNA was cloned unidirectionally into pSL1180. Inserts can be excised with EcoRI (5' end) and XhoI (3' end)."

## ORIGIN

Query Match 90.6%; Score 15.4; DB 8; Length 601;

Best Local Similarity 94.1%; Pred. No. 3.6e+03;

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17

Db 427 GCCCAGCGTTGGCCGAG 443

## RESULT 60

Cw480200

LOCUS

DEFINITION fbb001f239c09f0 Sorghum methylation filtered library (LibID: 104)  
Sorghum bicolor genomic clone fbb001f239c09, genomic survey sequence.

ACCESSION Cw480200

VERSION Cw480200.1

KEYWORDS GI:55228161

SOURCE GSS

ORGANISM Sorghum bicolor (sorghum)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD

clade; Panicoideae; Andropogoneae; Sorghum.

1 (bases 1 to 609)

REFERENCE Bedell, J.A., Budiman, M.A., Nunberg, A., Citek, R.W., Robbins, D.,

Jones, J., Flick, E., Rohlfing, T., Fries, J., Bradford, K.,

McMenamy, J., Smith, M., Holeman, H., Roe, B.A., Wiley, G., Korf, I.F.,

Rabinowicz, P.D., Lakey, N., McCombie, W.R., Jeddeloh, J.A. and

Martensen, R.A.

Sorghum genome sequencing by methylation filtration

PLoS Biol. 3 (1), e13 (2005)

15660154

COMMENT

Contact: Bedell JA

Orion Genomics, LLC

4041 Forest Park Ave, St. Louis, MO 63108, USA

Tel: 314 615 6979

Fax: 314 615 5975

Email: jbedell@oriongenomics.com

Plate: fbb001f239 row: c column: 09

Seq primer: f Forward

Class: methylation filtered

High quality sequence stop: 609.

## FEATURES

source

Location/Qualifiers

1..609

/organism="Sorghum bicolor"

/mol\_type="genomic DNA"

/cultivar="ATx623"

/db\_xref="taxon:4558"

/clone="fbb001f239c09"

104)"

/note="Organ: leaf; Vector: pBCSK(-); Site 1: HincII; DNA

prepared from purified nuclei was randomly-sheared,

end-repaired, size fractionated to enrich for the 0.5 to 5

kb fraction, ligated into HincII-digested pBCSK(-) vector

and electroporated into E. coli cells. This is a

methylation filtered library."

## ORIGIN

Query Match 90.6%; Score 15.4; DB 13; Length 609;

Best Local Similarity 94.1%; Pred. No. 3.6e+03;

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17

Db 437 GCCCAGCGTTGGCCGAG 453

## RESULT 61

BG929309

LOCUS

DEFINITION HNC60-1-H5.R HNC (Human Normal Cartilage) Homo sapiens cDNA, mRNA  
sequence.

ACCESSION BG929309

VERSION BG929309.1

KEYWORDS GI:14323832

SOURCE EST.

ORGANISM Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.

1 (bases 1 to 669)

REFERENCE Kumar, S., Connor, J.R., Dodds, R.A., Halsey, W., Van Horn, M., Mao, J.,

Sathre, G., Mui, P., Agarwal, P., Badger, A.M., Lee, J.C., Gowen, M. and

Lark, W.W.

Identification and initial characterization of 5000 expressed

sequenced tags (ESTs) each from adult human normal and

osteoarthritic cartilage cDNA libraries

Osteoarthr. Cartil. 9 (7), 641-653 (2001)

11597177

COMMENT

Contact: Sanjay Kumar

UW2109

GlaxoSmithKline

709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA

Tel: 610-270-7245

Fax: 610-270-5598

Email: sanjay.kumar-l@sk.com

Seq primer: T7.

Location/Qualifiers

1..669

source

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/tissue\_type="cartilage"

/lab\_host="E.coli DH10 B"

/clone\_lib="HNC (Human Normal Cartilage)"

/note="Vector: pSPORT 1; Site\_1: SalI; Site\_2: NotI;

Directional"

## ORIGIN

Query Match 90.6%; Score 15.4; DB 2; Length 669;

Best Local Similarity 94.1%; Pred. No. 3.7e+03;

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17

Db 559 GCCCAGCGTTGGCCGAG 575

```

RESULT 62
CW480201/c
LOCUS
DEFINITION
    CW480201
    671 bp DNA linear GSS 02-NOV-2004
    fbb001f239c09k0 Sorghum methylation filtered library (LibID: 104)
    Sorghum bicolor genomic clone fbb001f239c09, genomic survey
    sequence.
ACCESSION
    CW480201
    GI:55228162
VERSION
    GSS.
KEYWORDS
    Sorghum bicolor (sorghum)
SOURCE
    Sorghum bicolor
    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
    Spermatophyta; Magnoliopsida; Liliopsida; Poales; Poaceae; PACCAD
    clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE
    1 (bases 1 to 671)
    Bedell, J.A., Budiman, M.A., Nunberg, A., Citek, R.W., Robbins, D.,
    Jones, J., Flick, E., Rohlfing, T., Fries, J., Bradford, K.,
    McMenamy, J., Smith, M., Hleman, H., Roe, B.A., Wiley, G., Korf, I.F.,
    Rabinowicz, P.D., Lakey, N., McCombie, W.R., Jeddlo, J.A. and
    Martienssen, R.A.
    Sorghum genome sequencing by methylation filtration
    PLoS Biol. 3 (1), e13 (2005)
15660154
COMMENT
    Contact: Bedell JA
    Orion Genomics, LLC
    4041 Forest Park Ave, St. Louis, MO 63108, USA
    Tel: 314 615 6979
    Fax: 314 615 5975
    Email: jbedell@oriongenomics.com
    Plate: fbb001f239 row: c column: 09
    Seq primer: k Reverse
    Class: methylation filtered
    High quality sequence stop: 671.
FEATURES
    source
    1..671
    /organism="Sorghum bicolor"
    /mol_type="genomic DNA"
    /cultiivar="ATx623"
    /db_xref="taxon:4558"
    /clone_lib="fbb001f239c09"
    /clone_lib="Sorghum methylation filtered library (LibID:
    104)"
    /note="Organ: leaf; Vector: pCSK(-); Site_1: HincII; DNA
    prepared from purified nuclei was randomly sheared,
    end-repaired, size fractionated to enrich for the 0.5 to 5
    kb fraction, ligated into HincII-digested pCSK(-) vector
    and electroporated into E. coli cells. This is a
    methylation filtered library."
ORIGIN
    Query Match 90.6%; Score 15.4; DB 13; Length 671;
    Best Local Similarity 94.1%; Pred. No. 3.7e+03;
    Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
    Qy 1 GCCACGCTTGCCGCG 17
    Db 661 GCCACGCTTGCCGCG 645
    |||||||
RESULT 63
DN265926
LOCUS
DEFINITION
    DN265926
    689 bp mRNA linear EST 02-MAR-2005
    LIB30320_043_C10 SP6 1 LIB30320 Canis familiaris cDNA clone
    LIB30320_043_C10, mRNA sequence.
ACCESSION
    DN265926
    GI:60434536
VERSION
    EST.
KEYWORDS
    Canis familiaris (dog)
SOURCE
    Canis familiaris
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
    Canis.

```

```

REFERENCE
    1 (bases 1 to 689)
    Statten, N.R.
    Direct Submission (Statten, N.R.)
    JOURNAL
    Unpublished (2005)
    COMMENT
    Contact: Nick Statten
    Tel: 636 247 6855
    Email: nicholas.r.statten@pfizer.com.
FEATURES
    Location/Qualifiers
    source
    1..689
    /organism="Canis familiaris"
    /mol_type="mRNA"
    /db_xref="taxon:9615"
    /clone_lib="LIB30320_043_C10"
    /tissue_type="lymph node"
    /lab_host="DH10B"
    /clone_lib="LIB30320"
    /note="Vector: pCMV-Sport6; Site_1: NotI; Site_2: SalI;
    barcode 459"
ORIGIN
    Query Match 90.6%; Score 15.4; DB 9; Length 689;
    Best Local Similarity 94.1%; Pred. No. 3.7e+03;
    Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
    Qy 1 GCCACGCTTGCCGCG 17
    Db 612 GCCACGCTTGCCGCG 628
    |||||||
RESULT 64
DN441215
LOCUS
DEFINITION
    DN441215
    694 bp mRNA linear EST 08-MAR-2005
    LIB5338-114-Al-K2-B9 LIB5338 Canis familiaris cDNA clone
    CLN14242884, mRNA sequence.
ACCESSION
    DN441215
    GI:60637460
VERSION
    EST.
KEYWORDS
    Canis familiaris (dog)
SOURCE
    Canis familiaris
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
    Canis.
REFERENCE
    1 (bases 1 to 694)
    Statten, N.R.
    Direct Submission (Statten, N.R.)
    JOURNAL
    Unpublished (2005)
    COMMENT
    Contact: Nick Statten
    Tel: 636 247 6855
    Email: nicholas.r.statten@pfizer.com.
FEATURES
    Location/Qualifiers
    source
    1..694
    /organism="Canis familiaris"
    /mol_type="mRNA"
    /db_xref="taxon:9615"
    /clone_lib="CLN14242884"
    /tissue_type="kidney"
    /lab_host="DH10B"
    /clone_lib="LIB5338"
    /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; whole
    organ"
ORIGIN
    Query Match 90.6%; Score 15.4; DB 9; Length 694;
    Best Local Similarity 94.1%; Pred. No. 3.7e+03;
    Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
    Qy 1 GCCACGCTTGCCGCG 17
    Db 276 GCCACGCTTGCCGCG 292
    |||||||
RESULT 65
COL96932

```

LOCUS  
 DEFINITION GEO1\_2\_E06\_g1\_A029 Root gravitropism April 2003 test Pinus taeda linear mRNA EST 21-JUN-2004  
 ACCESSION cDNA clone GEO1\_2\_E06\_A029 5', mRNA sequence.  
 VERSION CO196932  
 KEYWORDS CO196932.1 GI:49008107  
 SOURCE EST.  
 ORGANISM Pinus taeda (loblolly pine)  
 Pinus taeda  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.  
 REFERENCE 1 (bases 1 to 721)  
 AUTHORS Pratt, L., Cordonnier-Pratt, M.-M., Lorenz, W.W., Zimmermann, C. and Dean, J.F.D.  
 TITLE An EST database from gravitropically stimulated loblolly pine (Pinus taeda) roots  
 JOURNAL Unpublished (2004)  
 COMMENT Other ESTs: GEO1\_2\_E06\_b1\_A029  
 Contact: Cordonnier-Pratt MM  
 Laboratory for Genomics and Bioinformatics  
 The University of Georgia, Department of Plant Biology  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 583 0210  
 Email: mmpratt@uga.edu  
 RNA prepared and library constructed by W. Walter Lorenz (School of Forest Resources, University of Georgia); plant material prepared by Craig Zimmermann (School of Forest Resources, University of Georgia) using rooted cuttings provided by the Forest Biology Research Cooperative (FBRC) and the CCLONES project at the University of Florida; sequencing done in the Laboratory for Genomics and Bioinformatics, University of Georgia. Sequence ends have been trimmed to exclude vector and regions below Phred quality 16. Three-prime sequences are presented as their reverse complement and have been trimmed to exclude polyA.  
 Seq primer: JENREV (CAGAACGCTATGACC).  
 FEATURES  
 source Location/Qualifiers  
 1..721  
 /organism="Pinus taeda"  
 /mol\_type="mRNA"  
 /strain="3 CCLONES"  
 /db\_xref="taxon:3352"  
 /clone="GEO1\_2\_E06\_A029"  
 /lab\_host="DH10B-T1 phage-resistant E. coli"  
 /clone\_lib="Root gravitropism April 2003 test"  
 /note="Vector: pSL1180; Site 1: EcoRI; Site 2: XhoI; The library was prepared from polyA+ RNA from the roots of 1-year-old loblolly pine (Pinus taeda) cuttings that were rooted and then planted in washed sand. The rooted cuttings were maintained for 27 days (April 2003 harvest) under ambient conditions in a local greenhouse. They were kept on a weekly regimen of 0.5x nutrient-complete Hoagland's solution and supplemented with additional water sufficient to maintain a 15% soil moisture content. Twenty-four hours (24 h) prior to harvesting roots for mRNA preparation, the potted trees were tipped 90 degree to lay the pots on their sides to induce a gravitropic response. Double-stranded cDNA was cloned unidirectionally into pSL1180. Inserts can be excised with EcoRI (5' end) and XhoI (3' end)."  
 ORIGIN  
 Query Match 90.6%; Score 15.4; DB 8; Length 721;  
 Best Local Similarity 94.1%; Pred. No. 3.7e+03;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Oy 1 GCCCAGCTTGGCCGAG 17  
 |||||  
 Db 567 GCCCAGCTTGGCCGAG 583  
 RESULT 66  
 CC111772/c  
 LOCUS CC111772 linear DNA 744 bp GSS 16-APR-2003

DEFINITION NDL\_46M15.T7 Notre Dame Liverpool Aedes aegypti genomic clone  
 ACCESSION NDL\_46M15, genomic survey sequence.  
 VERSION CC111772  
 KEYWORDS CC111772.1 GI:29980827  
 SOURCE GSS.  
 ORGANISM Aedes aegypti (yellow fever mosquito)  
 Aedes aegypti  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Culicidae; Culicinae; Culicini; Aedes; Stegomyia.  
 REFERENCE 1 (bases 1 to 744)  
 AUTHORS Loftus, B., Shetty, J., Knudson, D. and Severson, D.  
 TITLE BAC end sequencing of Aedes aegypti  
 JOURNAL Unpublished (2003)  
 COMMENT Other GSSs: NDL\_46M15.SP6  
 Contact: Brendan Loftus  
 Department of Eukaryotic Genomics  
 TIGR  
 9712 Medical Center Drive, Rockville, MD 20850, USA  
 Tel: 301-838-3543  
 Fax: 301-838-0208  
 Email: enta@tigr.org  
 Library was provided by David Severson  
 Seq primer: T7  
 Class: BAC ends.  
 FEATURES  
 source Location/Qualifiers  
 1..744  
 /organism="Aedes aegypti"  
 /mol\_type="genomic DNA"  
 /strain="Liverpool"  
 /db\_xref="taxon:7159"  
 /clone="NDL\_46M15"  
 /clone\_lib="Notre Dame Liverpool"  
 /note="Vector: pSCBAC1; Site 1: Hind III; The library was prepared from whole body tissue of newly hatched L1 larvae by David Severson at the University of Notre Dame and Hongbin Zhang"  
 ORIGIN  
 Query Match 90.6%; Score 15.4; DB 12; Length 744;  
 Best Local Similarity 94.1%; Pred. No. 3.7e+03;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Oy 1 GCCCAGCTTGGCCGAG 17  
 |||||  
 Db 138 GCCCAGCTTGGCCGAG 122  
 RESULT 67  
 CW576188  
 LOCUS CW576188 749 bp DNA linear GSS 22-OCT-2004  
 DEFINITION OA\_ABA0104L02.f OA\_ABA Oryza australiensis genomic clone  
 ACCESSION OA\_ABA0104L02 5', genomic survey sequence.  
 VERSION CW576188  
 KEYWORDS CW576188.1 GI:54458858  
 SOURCE GSS.  
 ORGANISM Oryza australiensis  
 Oryza australiensis  
 Eukaryota; Viridiplantae; Streptophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP clade; Ehrhartoideae; Oryzaceae; Oryza.  
 REFERENCE 1 (bases 1 to 749)  
 AUTHORS Kim, H., Yu, Y., Stum, D., Yost, D., Rao, K., Luo, M., Jetty, R., Kudrna, D., Muller, C., Hatfield, J., Soderlund, C. and Wing, R.  
 TITLE OMAP Project  
 JOURNAL Unpublished (2004)  
 COMMENT Contact: Rod A. Wing  
 Arizona Genomics Institute  
 University of Arizona  
 Forbes Building Room 303, Tucson, AZ 85721-0036, USA  
 Tel: 520 626 9595  
 Fax: 520 621 1259  
 Email: rwing@genome.arizona.edu

## PCR Primers

FORWARD: TAA TAC GAC TCA CTA TAG GG  
 BACKWARD: CAC TCA TTA GGC ACC CCA  
 Plate: 0104 row: 1 column: 02  
 Seq primer: TAA TAC GAC TCA CTA TAG GG  
 Class: BAC ends.

## FEATURES

Location/Qualifiers  
 1. .749  
 /organism="Oryza australiensis"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:4532"  
 /clone="OA\_Aba0104L02"  
 /tissue\_type="young leaves"  
 /lab\_host="DH10B T1 phage resistant"  
 /clone\_lib="OA\_ABA"  
 /note="Vector: pAGIBAC1; Site\_1: HindIII; Site\_2: HindIII"

## ORIGIN

Query Match 90.6%; Score 15.4; DB 13; Length 749;  
 Best Local Similarity 94.1%; Pred. No. 3.7e+03;  
 Mismatches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCACGCTGGCGGAG 17

Db 379 GCTCACGCTGGCGGAG 395

## RESULT 68

## LOCUS

CR374717 788 bp mRNA linear EST 22-APR-2004  
 CR374717 AGENAE Rainbow trout normalized testis library (tcbl)  
 Oncorhynchus mykiss cDNA clone tcb10013c.f.10 5prim, mRNA sequence.

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

Oncorhynchus mykiss (rainbow trout)

Oncorhynchus mykiss

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Euteleostei;

Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.

1 (bases 1 to 788)

Govoroun, M., Guiguen, Y. and Le Gac, F.

Construction and primary characterization of normalized cDNA

libraries in rainbow trout, *Oncorhynchus mykiss*

Unpublished (2003)

Contact: Guiguen Y

INRA - SCRIBE

Campus de Beaulieu, RENNES cedex, 35042, France

Tel: 02.23.48.50.09

Fax: 02.23.48.50.20

Email: Yann.Guiguen@beaulieu.rennes.inra.fr

Sequence cleaned of vector, adaptor and repetitions. Contact us

at [sigenasupport@jouy.inra.fr](mailto:sigenasupport@jouy.inra.fr) to obtain the chromatogram of this

sequence.

Plate: 0013 row: f column: 10

Seq primer: M13R.

Location/Qualifiers

1. .788

/organism="Oncorhynchus mykiss"

/mol\_type="mRNA"

/db\_xref="taxon:8022"

/clone="tcb10013c.f.10"

/tissue\_type="testis"

/lab\_host="DH10B"

/clone\_lib="AGENAE Rainbow trout normalized testis library

(tcbl)"

/note="Vector: p7T3D-PacI; Clone distribution : AGENAE

Resource centre, Francois PIUMI,

Francois.Piumi@jouy.inra.fr, INRA, CEA Radiobiologie et

Etude du genome (LURE), Domaine de Vilvert, 78352,

Jouy-en-Josas cedex, FRANCE, +33 (0) 1.34.65.28.02, +33

(0) 1.34.65.22.73"

## ORIGIN

Query Match 90.6%; Score 15.4; DB 8; Length 788;  
 Best Local Similarity 94.1%; Pred. No. 3.7e+03;  
 Mismatches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCACGCTGGCGGAG 17

Db 341 GCCCAAGCTGGCCGAG 357

## RESULT 69

## LOCUS

BF700990/c 797 bp mRNA linear EST 22-DEC-2000  
 602127938F1 NIH\_MGC\_56 Homo sapiens cDNA clone IMAGE:4284825 5',  
 mRNA sequence.

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.

1 (bases 1 to 797)

NIH-MGC <http://mgs.nci.nih.gov/>

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: [cgabs-y@mail.nih.gov](mailto:cgabs-y@mail.nih.gov)

Tissue Procurement: ATCC

cDNA Library Preparation: CLONTECH Laboratories, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLCm118 row: 1 column: 10

High quality sequence start: 4

High quality sequence stop: 500.

Location/Qualifiers

1. .797

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:4284825"

/tissue\_type="primitive neuroectoderm"

/lab\_host="DH10B (T1 phage-resistant)"

/clone\_lib="NIH\_MGC 56"

/note="Organ: brain; Vector: pDNR-LTB (Clontech); Site\_1:

SfiI (ggccattatggcc); Site\_2: SfiI (ggccattatggcc);

Double-stranded cDNA was prepared from cell line RNA. 5'

and 3' adaptors were used in cloning as follows: 5'

adaptor sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor

sequence: 5'-ATTCTAGAGCGGCGGCGGACATG-dT(30)BN-3',

(where B = A, C, or G and N = A, C, G, or T). Average

insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies

contained inserts by PCR. This library was enriched for

full-length clones and was constructed by Clontech

Laboratories (Palo Alto, CA)."

## ORIGIN

Query Match 90.6%; Score 15.4; DB 7; Length 797;  
 Best Local Similarity 94.1%; Pred. No. 3.7e+03;  
 Mismatches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCACGCTGGCGGAG 17

Db 673 GCCACGCTGGCCAG 657

## RESULT 70

## LOCUS

CO196873 798 bp mRNA linear EST 21-JUN-2004

**DEFINITION** GE01\_2\_E06.b1 A029 Root gravitropism April 2003 test Pinus taeda cDNA clone GE01\_2\_E06\_A029 3', mRNA sequence.  
**ACCESSION** CO196873  
**VERSION** CO196873.1 GI:49008048  
**KEYWORDS** EST.  
**SOURCE** Pinus taeda (loblolly pine)  
**ORGANISM** Pinus taeda  
**REFERENCE** Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus. 1 (bases 1 to 798)  
**AUTHORS** Pratt,L., Cordonnier-Pratt,M.-M., Lorenz,W.W., Zimmermann,C. and Dean,J.F.D.  
**TITLE** An EST database from gravitropically stimulated loblolly pine (Pinus taeda) roots  
**JOURNAL** Unpublished (2004)  
**COMMENT** Other ESTs: GE01\_2\_E06.g1\_A029  
 Contact: Cordonnier-Pratt MM  
 Laboratory for Genomics and Bioinformatics  
 The University of Georgia, Department of Plant Biology  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 583 0210  
 Email: mmpratt@uga.edu  
 RNA prepared and library constructed by W. Walter Lorenz (School of Forest Resources, University of Georgia); plant material prepared by Craig Zimmermann (School of Forest Resources, University of Georgia) using rooted cuttings provided by the Forest Biology Research Cooperative (FBRC) and the CCLONES project at the University of Florida; sequencing done in the Laboratory for Genomics and Bioinformatics, University of Georgia. Sequence ends have been trimmed to exclude vector and regions below Phred quality 16. Three-prime sequences are presented as their reverse complement and have been trimmed to exclude polyA.  
 Seq primer: M13-21 (TGTAACACGACGCCAGT)  
 POLYA=No.

# **FEATURES** source

1. 798  
 /organism="Pinus taeda"  
 /mol\_type="mRNA"  
 /strain="3 CCLONES"  
 /db\_xref="taxon:3352"  
 /clone="GE01\_2\_E06\_A029"  
 /lab\_host="DH10B-T1 phage-resistant E. coli"  
 /clone\_lib="Root gravitropism April 2003 test"  
 /note="vector: pSL180; Site 1: EcoRI; Site 2: XhoI; The library was prepared from polyA+ RNA from the roots of 1-year-old loblolly pine (Pinus taeda) cuttings that were rooted and then planted in washed sand. The rooted cuttings were maintained for 27 days (April 2003 harvest) under ambient conditions in a local greenhouse. They were kept on a weekly regimen of 0.5x nutrient-complete Hoagland's solution and supplemented with additional water sufficient to maintain a 15% soil moisture content. Twenty-four hours (24 h) prior to harvesting roots for mRNA preparation, the potted trees were tipped 90 degree to lay the pots on their sides to induce a gravitropic response. Double-stranded cDNA was cloned unidirectionally into pSL180. Inserts can be excised with EcoRI (5' end) and XhoI (3' end)."  
 Location/Qualifiers

## **ORIGIN**

Query Match 90.6%; Score 15.4; DB 8; Length 798;  
 Best Local Similarity 94.1%; Pred. No. 3.7e+03;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGCCGAG 17

|||||

Db 578 GCCCAGCTTGCCGAG 594

RESULT 71

BZ455483/c

LOCUS BZ455483 802 bp DNA linear GSS 13-DEC-2002

**DEFINITION**  
**ACCESSION**  
**VERSION**  
**KEYWORDS**  
**SOURCE**  
**ORGANISM**

BONIS66TR BO 1.6 2 KB tot Brassica oleracea genomic clone BONIS66, genomic survey sequence.  
 BZ455483  
 BZ455483.1 GI:26732690  
 GSS.  
 Brassica oleracea

**REFERENCE**  
**AUTHORS**

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.

**TITLE**

1 (bases 1 to 802)  
 Ayele,M., Haas,B.J., Kumar,N., Wu,H., Xiao,Y., Van Aken,S., Uterback,T.R., Wortman,J.R., White,O.R. and Town,C.D.  
 Whole genome shotgun sequencing of Brassica oleracea and its application to gene discovery and annotation in Arabidopsis

**JOURNAL**  
**PUBMED**  
**COMMENT**

Genome Res. 15 (4), 487-495 (2005)  
 15805490  
 Other GSSs: BONIS66TF  
 Contact: Chris Town

**FEATURES**  
 source

1. 802  
 /organism="Brassica oleracea"  
 /mol\_type="genomic DNA"  
 /strain="T01000H3"  
 /db\_xref="taxon:3712"  
 /clone="BONIS66"  
 /clone\_lib="BO 1.6 2 KB tot"  
 /note="vector: pHOS1; Site 1: BstXI; 1.6-2 kb sheared total DNA inserted into pHOS1 using BstXI linkers"

Location/Qualifiers

1. 802  
 /organism="Brassica oleracea"  
 /mol\_type="genomic DNA"  
 /strain="T01000H3"  
 /db\_xref="taxon:3712"  
 /clone="BONIS66"  
 /clone\_lib="BO 1.6 2 KB tot"  
 /note="vector: pHOS1; Site 1: BstXI; 1.6-2 kb sheared total DNA inserted into pHOS1 using BstXI linkers"

**ORIGIN**

Query Match 90.6%; Score 15.4; DB 11; Length 802;  
 Best Local Similarity 94.1%; Pred. No. 3.7e+03;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGCCGAG 17

|||||

Db 741 GCCCAGCTTGCCGAG 725

RESULT 72

LOCUS

BONIS66TR BO 1.6 2 KB tot Brassica oleracea genomic clone  
 BZ455483  
 BZ455483.1 GI:26732690  
 GSS.  
 Brassica oleracea

**REFERENCE**  
**AUTHORS**

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP clade; Ehrhartoideae; Oryzaceae; Oryza.

**TITLE**

**JOURNAL**  
**COMMENT**

1 (bases 1 to 816)  
 Kim,H., Yu,X., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R., Kudrna,D., Muller,C., Hattfield,J., Soderlund,C. and Wing,R.  
 OMAF Project  
 Unpublished (2004)  
 Contact: Rod A. Wing  
 Arizona Genomics Institute  
 University of Arizona  
 Forbes Building Room 303, Tucson, AZ 85721-0036, USA  
 Tel: 520 626 9595  
 Fax: 520 621 1259  
 Email: rwing@genome.arizona.edu

**ORIGIN**

Query Match 90.6%; Score 15.4; DB 11; Length 802;  
 Best Local Similarity 94.1%; Pred. No. 3.7e+03;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGCCGAG 17

|||||

Db 741 GCCCAGCTTGCCGAG 725

RESULT 72

LOCUS

BONIS66TR BO 1.6 2 KB tot Brassica oleracea genomic clone  
 BZ455483  
 BZ455483.1 GI:26732690  
 GSS.  
 Brassica oleracea

**REFERENCE**  
**AUTHORS**

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP clade; Ehrhartoideae; Oryzaceae; Oryza.

**TITLE**

**JOURNAL**  
**COMMENT**

1 (bases 1 to 816)  
 Kim,H., Yu,X., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R., Kudrna,D., Muller,C., Hattfield,J., Soderlund,C. and Wing,R.  
 OMAF Project  
 Unpublished (2004)  
 Contact: Rod A. Wing  
 Arizona Genomics Institute  
 University of Arizona  
 Forbes Building Room 303, Tucson, AZ 85721-0036, USA  
 Tel: 520 626 9595  
 Fax: 520 621 1259  
 Email: rwing@genome.arizona.edu

**ORIGIN**

Query Match 90.6%; Score 15.4; DB 8; Length 798;  
 Best Local Similarity 94.1%; Pred. No. 3.7e+03;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGCCGAG 17

|||||

Db 578 GCCCAGCTTGCCGAG 594

RESULT 71

BZ455483/c

LOCUS BZ455483 802 bp DNA linear GSS 13-DEC-2002

QY	1	GCCACGCTTGCCGAG	17
DB	379	GCTCAGCTTGCCGAG	395
RESULT 74			
CA495187/c			
LOCUS	CA495187/c		
DEFINITION			
ACCESSION	CA495187		
VERSION	CA495187.1	GI:24958247	
KEYWORDS	EST.		
SOURCE	Mus musculus	(house mouse)	
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;		
TITLE	Sciurognathi; Murioidea; Muridae; Murinae; Mus.		
JOURNAL	1 (bases 1 to 851)		
COMMENT	NIH-MGC http://mgc.nci.nih.gov/ National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapsb@mail.nih.gov Tissue Procurement: Bradfield Laboratory cdNA Library Preparation: Mark Bittinger CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LKX150 row: m column: 04 High quality sequence stop: 230. Location/Qualifiers 1..851 /organism="Mus musculus" /mol_type="mRNA" /db_xref="taxon:10090" /clone="IMAGE:6776333" /lab_host="DH10B (phage-resistant)" /clone_lib="NIH_MGC_154" /note="Organ: Kidney; TCDD (DMSO vehicle) Treated 48 Hours IP injections; Vector: pDONR201; Site_1: attP2; Site_2: attP1; cDNA made by oligo-dt with attB2 site and directionally cloned. Priming sequence: 5'-TTTCGTGCGAGCGGCCACCATTGTACAGAAGCTGGGTTTTTTTTTTT TTTT-3'. Full-length enriched library was constructed using the GeneRacer kit by Invitrogen, library amplification 16 cycles. Library constructed by Mark Bittinger in the Bradfield laboratory (McArdle Laboratory for Cancer Research, University of Wisconsin). Note: this is a NIH_MGC Library."		
FEATURES			
source			
ORIGIN			
Query Match	90.6%;	Score 15.4;	DB 4; Length 851;
Best Local Similarity	94.4%;	Pred. No. 3.7e+03;	
Matches	16; Conservative	0; Mismatches	1; Indels 0; Gaps 0;
QY	1	GCCACGCTTGCCGAG	17
DB	347	GCCACGCTTGCCGAG	331
RESULT 75			
CC112512/c			
LOCUS	CC112512/c		
DEFINITION	NDL.37C12.sp6 Notre Dame Liverpool Aedes aegypti genomic clone		
ACCESSION	CC112512		
VERSION	CC112512.1	GI:29991567	
KEYWORDS	Aedes aegypti (yellow fever mosquito)		
SOURCE			

```

ORGANISM
Aedes aegypti
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea,
Culicidae; Culicinae; Culicini; Aedes; Stegomyia.
1 (bases 1 to 899)
REFERENCE
AUTHORS Loftus,B., Shetty,J., Knudson,D. and Severson,D.
TITLE BAC end sequencing of Aedes aegypti
JOURNAL Unpublished (2003)
COMMENT Other GSSs: NDL.37C12.T7
Contact: Brendan Loftus
Department of Eukaryotic Genomics
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-3543
Fax: 301-838-0208
Email: enta@tigr.org
Library was provided by David Severson
Seq primer: SP6
Class: BAC ends.
FEATURES
source
Location/Qualifiers
1..899
/organism="Aedes aegypti"
/mol_type="genomic DNA"
/db_xref="taxon:7159"
/clone="NDL.37C12"
/note="Vector: pECBAC1; Site_1: Hind III; The library was
prepared from whole body tissue of newly hatched L1 larvae
by David Severson at the University of Notre Dame and
Hongbin Zhang"
ORIGIN
Query Match 90.6%; Score 15.4; DB 12; Length 899;
Best Local Similarity 94.1%; Pred. No. 3.7e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 GCCCAGCTTGCCGAG 17
|||||
Db 152 GCCCAGCATGCCGAG 136
|||||

RESULT 76
CC851671/c
LOCUS CC851671 914 bp DNA linear GSS 24-JUL-2003
DEFINITION NDL.55M12.SP6 Notre Dame Liverpool Aedes aegypti genomic clone
Notrebame Liverpool-55M12, genomic survey sequence.
ACCESSION CC851671
VERSION CC851671.1 GI:33203800
KEYWORDS GSS.
SOURCE Aedes aegypti (yellow fever mosquito)
ORGANISM
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Culicidae; Culicinae; Culicini; Aedes; Stegomyia.
1 (bases 1 to 914)
REFERENCE
AUTHORS Loftus,B., Shetty,J., Knudson,D. and Severson,D.
TITLE BAC end sequencing of Aedes aegypti
JOURNAL Unpublished (2003)
COMMENT Other GSSs: NDL.55M12.T7
Contact: Brendan Loftus
Department of Eukaryotic Genomics
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-3543
Fax: 301-838-0208
Email: enta@tigr.org
Library was provided by David Severson
Seq primer: SP6
Class: BAC ends.
FEATURES
source
Location/Qualifiers
1..914
/organism="Aedes aegypti"

```

```

/mol_type="genomic DNA"
/strain="Liverpool"
/db_xref="taxon:7159"
/clone="Notrebame Liverpool-55M12"
/clone_lib="Notre Dame Liverpool"
/note="Vector: pECBAC1; Site_1: Hind III; The library was
prepared from whole body tissue of newly hatched L1 larvae
by David Severson at the University of Notre Dame and
Hongbin Zhang"
ORIGIN
Query Match 90.6%; Score 15.4; DB 12; Length 914;
Best Local Similarity 94.1%; Pred. No. 3.7e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 GCCCAGCTTGCCGAG 17
|||||
Db 189 GCCCAGCATGCCGAG 173
|||||

RESULT 77
CNS02J9X
LOCUS CNS02J9X 929 bp DNA linear GSS 01-SEP-2000
DEFINITION Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
143K05 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL199950
VERSION AL199950.1 GI:7838101
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontoidea; Tetraodontidae; Tetraodon.
1
Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Estimate of human gene number provided by genome-wide analysis
using Tetraodon nigroviridis DNA sequence
Nat. Genet. 25 (2), 235-238 (2000)
10835645
2
Roest Crolius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
Saurin,W., Bernot,A. and Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Genome Res. 10 (7), 939-949 (2000)
10899143
3 (bases 1 to 929)
Direct Submission
Genoscope.
Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr
- Web : www.genoscope.cns.fr)
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/tetraodon.
FEATURES
source
Location/Qualifiers
1..929
/organism="Tetraodon nigroviridis"
/mol_type="genomic DNA"
/db_xref="taxon:99883"
/clone="143K05"
/clone_lib="G"
/note="Genoscope sequence ID : CONG143AF03SP1
end : PUC-Ori"
ORIGIN
Query Match 90.6%; Score 15.4; DB 14; Length 929;

```

```

Best Local Similarity 94.1%; Pred. No. 3.7e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCCAGCGCTGGCGGAG 17
|||||
Db 448 GCCCAGCGCTGGCGGAG 464

RESULT 78
BO685729 940 bp mRNA linear EST 15-JUL-2002
LOCUS AGENCOURT_8344399 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6251057
DEFINITION 5', mRNA sequence.
ACCESSION BO685729
VERSION BO685729.1 GI:21811045
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 940)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
plate: LICM2395 row: b column: 18
High quality sequence stop: 538.
Location/Qualifiers
1. .940
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6251057"
/tissue_type="ductal carcinoma, cell line"
/lab_hosts="DH10B (phage-resistant)"
/clone_lib="NIH MGC 110"
/site="Organ: pancreas; Vector: pOTB7; Site 1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAGGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

ORIGIN
Query Match 90.6%; Score 15.4; DB 3; Length 940;
Best Local Similarity 94.1%; Pred. No. 3.7e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCCAGCGCTGGCGGAG 17
|||||
Db 687 GCCCAGCGCTGGCGGAG 703

RESULT 79
DW341470 1058 bp mRNA linear EST 09-JAN-2006
LOCUS DW341470
DEFINITION PP_LEC0004H16f Peach shoot Prunus persica cDNA clone
ACCESSION DW341470
VERSION DW341470.1 GI:84627860
KEYWORDS EST.
SOURCE Prunus persica (peach)

```

```

Prunus persica
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids I; Rosales; Rosaceae; Amygdaloideae; Prunus.
1 (bases 1 to 1058)
Abbott, A., Zhebentyayeva, T., Main, D., Jung, S., Staton, M.,
Jesudurai, C. and Wing, R.
Peach Model Genome for Rosaceae Shoot ESTs
Unpublished (2006)
Contact: Abbott, A.
Dept of Genetics and Biochemistry
Clemson University
122 Long Hall, Clemson University, Clemson, SC 29634, USA
Tel: 864 656 3060
Fax: 864 656 6879
Email: aalbert@clemson.edu
Total High Quality bases = 109
Seq primer: TAATACGACTCACTATAGG
High quality sequence stop: 1058.
Location/Qualifiers
1. .1058
/organism="Prunus persica"
/mol_type="mRNA"
/cultivar="Loring"
/db_xref="taxon:3760"
/clone="PP_LEC0004H16f"
/tissue_type="Shoot"
/lab_host="E. coli"
/clone_lib="Peach shoot"
/notes="Vector: pBluescript II SK(-); Site_1: EcoRI;
Site_2: XhoI"

ORIGIN
Query Match 90.6%; Score 15.4; DB 10; Length 1058;
Best Local Similarity 94.1%; Pred. No. 3.7e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCCAGCGTTGGCGGAG 17
|||||
Db 968 GCCCATGCTTGGCGGAG 984

RESULT 80
DRI25588/c 1117 bp mRNA linear EST 15-JUN-2005
LOCUS 49128899 Drosophila pseudoobscura embryonic cDNA library Drosophila
DEFINITION pseudoobscura cDNA clone G11 5', mRNA sequence.
ACCESSION DRI25588
VERSION DRI25588.1 GI:67842286
KEYWORDS EST.
SOURCE Drosophila pseudoobscura
ORGANISM Drosophila pseudoobscura
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1117)
Richards, S., Liu, Y., Bettencourt, B.R., Hradecky, P., Letovsky, S.,
Nielsen, R., Thornton, K., Hubisz, M.J., Chen, R., Meisel, R.P.,
Couronne, O., Hua, S., Smith, M.A., Zhang, P., Liu, J., Bussemaker, H.J.,
van Batenburg, M.F., Howells, S.L., Scherer, S.E., Sodergren, E.,
Matthews, B.B., Crosby, M.A., Schroeder, A.J., Ortiz-Barrientos, D.,
Rives, C.M., Metzker, M.L., Muzny, D.M., Scott, G., Steffen, D.,
Wheeler, D.A., Worley, K.C., Havlak, P., Durbin, K.J., Sgan, A.,
Gill, R., Hume, J., Morgan, M.B., Miner, G., Hamilton, C., Huang, Y.,
Waldron, L., Verduzco, D., Clerc-Blankenburg, K.P., Dubchak, I.,
Noor, M.A., Anderson, W., White, K.P., Clark, A.G., Schaeffer, S.W.,
Gelbart, W., Weinstock, G.M. and Gibbs, R.A.
Comparative genome sequencing of Drosophila pseudoobscura:
chromosomal, gene, and cis-element evolution
Genome Res. 15 (1), 1-18 (2005)
15632085
Contact: Stephen Richards
Human Genome Sequencing Center

```



Baylor College of Medicine  
One Baylor Plaza, Houston, TX 77030, USA  
Tel: 713-798-6667  
Email: stephen@bcm.tmc.edu  
NCBI Trace Archive: 226701216  
Insert Length: 1750 Std Error: 0.25.

# FEATURES

Location/Qualifiers  
1. .1117  
/organism="Drosophila pseudoobscura"  
/mol\_type="mRNA"  
/db\_xref="taxon:7237"  
/clone="G11"  
/dev\_stage="0-18h embryos"  
/clone\_lib="Drosophila pseudoobscura embryonic cDNA library"  
/note="Vector: pOTB7; Site 1: EcoRI; Site 2: XhoI; oligo dt priming from poly A+ RNA, directionally cloned"

## ORIGIN

Query Match 90.6%; Score 15.4; DB 9; Length 1117;  
Best Local Similarity 94.1%; Pred. No. 3.7e+03;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 GCCACGCTTGCCGAG 17  
||| ||||| ||||| |||||  
Db 1009 GCTCAGCTTGCCGAG 993

## RESULT 81

BE729821/c  
LOCUS BE729821 1156 bp mRNA linear EST 15-SEP-2000  
DEFINITION 60156473F1 NIH\_MGC\_20 Homo sapiens cDNA clone IMAGE:3926001 5', mRNA sequence.  
ACCESSION BE729821 GI:10143813  
VERSION BE729821.1  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1156)  
NIH-MGC http://mgc.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: ATCC/DCTP/DTP  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
Plate: LICM753 row: e column: 10  
High quality sequence start: 16  
High quality sequence stop: 545.

## FEATURES

Location/Qualifiers  
1. .1156  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3926001"  
/tissue\_type="melanotic melanoma"  
/lab\_host="NIH\_MGC\_20"  
/clone\_lib="NIH\_MGC\_20"  
/note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

## ORIGIN

Query Match 90.6%; Score 15.4; DB 7; Length 1156;  
Best Local Similarity 94.1%; Pred. No. 3.7e+03;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 GCCACGCTTGCCGAG 17  
||| ||||| ||||| |||||  
Db 96 GCCCATGCTTGCCGAG 80

## RESULT 82

CD256036  
LOCUS CD256036 1263 bp mRNA linear EST 22-MAY-2003  
DEFINITION AGENCOURT 14159547 NICHD\_XGC Brnl Xenopus laevis cDNA clone IMAGE:6952009 5', mRNA sequence.  
ACCESSION CD256036  
VERSION CD256036.1  
KEYWORDS GI:31016502  
SOURCE EST.  
ORGANISM Xenopus laevis (African clawed frog)  
Xenopus laevis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae; Xenopodinae; Xenopus; Xenopus.  
REFERENCE 1 (bases 1 to 1263)  
NIH-MGC http://mgc.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Daniela S. Gerhardt, Ph.D.  
Office of Cancer Genomics  
National Cancer Institute / NIH  
Bldg. 31 Rm10A07 Bethesda, MD 20892  
Email: cgabbs-remail.nih.gov  
Tissue Procurement:  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov  
Plate: LHAM14571 row: p column: 24  
High quality sequence stop: 425.

## FEATURES

Location/Qualifiers  
1. .1263  
/organism="Xenopus laevis"  
/mol\_type="mRNA"  
/db\_xref="taxon:8355"  
/clone="IMAGE:6952009"  
/dev\_stage="adult"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NICHD\_XGC\_Brnl"  
/note="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.5 kb. Constructed by Life Technologies. Note: This is a Xenopus Gene Collection (XGC) library."

## ORIGIN

Query Match 90.6%; Score 15.4; DB 5; Length 1263;  
Best Local Similarity 94.1%; Pred. No. 3.7e+03;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 GCCACGCTTGCCGAG 17  
||| ||||| ||||| |||||  
Db 620 GCCCATCTTGCCGAG 636

## RESULT 83

CNS05ZL3  
LOCUS CNS05ZL3 1388 bp mRNA linear HTC 05-JUL-2005  
DEFINITION Tetraodon nigroviridis full-length cDNA.  
ACCESSION CNS05ZL3  
VERSION CNS05ZL3.2 GI:56244151

```

KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS

1
Jailion,O., Aury,J.M., Brunet,F., Petit,J.L., Stange-Thomann,N.,
Mauceli,E., Bouneau,L., Fischer,C., Ozouf-Costaz,C., Bernot,A.,
Nicaud,S., Jaffe,D., Fisher,S., Lutfalla,G., Dossat,C.,
Segurens,B., Dasilva,C., Salanoubat,M., Levy,M., Boudet,N.,
Castellano,S., Anthouard,V., Jubin,C., Castellio,L., Poulain,J., De
Vacherie,B., Biemont,C., Skalli,Z., Cattolico,L., Coutanceau,J.P.,
Berardinis,V., Cruaud,C., Duprat,S., Brottier,P., Coutanceau,J.P.,
Gouzy,J., Parra,G., Lardier,G., Chapple,C., McKernan,K.J.,
McEwan,P., Bosak,S., Kellis,M., Volff,J.N., Guigo,R., Zody,M.C.,
Mesirov,J., Lindblad-Toh,K., Laudet,V., Schachter,V., Quetier,F.,
Robinson-Rechavi,M., Laudet,V., Schachter,V., Quetier,F.,
Saurin,W., ScarPELLI,C., Wincker,P., Lander,E.S., Weissbach,J.,
and Roest Crolius,H.
Genome duplication in the teleost fish Tetraodon nigroviridis
reveals the early vertebrate proto-karyotype
Nature 431 (7011), 946-957 (2004)
15496914
PUBMED
2 (bases 1 to 1388)
REFERENCE
AUTHORS

Direct Submission
Submitted (25-NOV-2004) Genoscope - Centre National de Sequencage -
: 2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
(E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
On Dec 3, 2004 this sequence version replaced gi:51145246.
The sequences are based on single pass reads.
More information available at
http://www.genoscope.cns.fr/tetraodon.
Location/Qualifiers
1. .1388
/organism="Tetraodon nigroviridis"
/mol_type="mRNA"
/db_xref="taxon:99883"
/tissue_type="Liver"

ORIGIN

Query Match 90.6%; Score 15.4; DB 6; Length 1388;
Best Local Similarity 94.1%; Pred. No. 3.7e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCACGCTGGCCGAG 17
|||||||
Db 707 GCCACGCTAGCCGAG 723

RESULT 84
CNSOGOND 1454 bp mRNA linear HTC 05-JUL-2005
LOCUS
DEFINITION Tetraodon nigroviridis full-length cDNA.
ACCESSION CR696861
VERSION CR696861.2 GI:56287441
KEYWORDS
SOURCE
ORGANISM

Htc; cDNA; full-length; full-length cDNA; Tetraodon nigroviridis.

REFERENCE
AUTHORS

1
Jailion,O., Aury,J.M., Brunet,F., Petit,J.L., Stange-Thomann,N.,
Mauceli,E., Bouneau,L., Fischer,C., Ozouf-Costaz,C., Bernot,A.,
Nicaud,S., Jaffe,D., Fisher,S., Lutfalla,G., Dossat,C.,
Segurens,B., Dasilva,C., Salanoubat,M., Levy,M., Boudet,N.,
Castellano,S., Anthouard,V., Jubin,C., Castellio,L., Poulain,J., De
Vacherie,B., Biemont,C., Skalli,Z., Cattolico,L., Coutanceau,J.P.,
Berardinis,V., Cruaud,C., Duprat,S., Brottier,P., Coutanceau,J.P.,
Gouzy,J., Parra,G., Lardier,G., Chapple,C., McKernan,K.J.,
McEwan,P., Bosak,S., Kellis,M., Volff,J.N., Guigo,R., Zody,M.C.,
Mesirov,J., Lindblad-Toh,K., Laudet,V., Schachter,V., Quetier,F.,
Robinson-Rechavi,M., Laudet,V., Schachter,V., Quetier,F.,
Saurin,W., ScarPELLI,C., Wincker,P., Lander,E.S., Weissbach,J.,
and Roest Crolius,H.
Genome duplication in the teleost fish Tetraodon nigroviridis
reveals the early vertebrate proto-karyotype
Nature 431 (7011), 946-957 (2004)
15496914
PUBMED
2 (bases 1 to 1388)
REFERENCE
AUTHORS

Direct Submission
Submitted (25-NOV-2004) Genoscope - Centre National de Sequencage -
: 2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
(E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
On Dec 3, 2004 this sequence version replaced gi:51145246.
The sequences are based on single pass reads.
More information available at
http://www.genoscope.cns.fr/tetraodon.
Location/Qualifiers
1. .1388
/organism="Tetraodon nigroviridis"
/mol_type="mRNA"
/db_xref="taxon:99883"
/tissue_type="Liver"

ORIGIN

Query Match 90.6%; Score 15.4; DB 6; Length 1388;
Best Local Similarity 94.1%; Pred. No. 3.7e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCACGCTGGCCGAG 17
|||||||
Db 707 GCCACGCTAGCCGAG 723

RESULT 84
CNSOGOND 1454 bp mRNA linear HTC 05-JUL-2005
LOCUS
DEFINITION Tetraodon nigroviridis full-length cDNA.
ACCESSION CR696861
VERSION CR696861.2 GI:56287441
KEYWORDS
SOURCE
ORGANISM

Htc; cDNA; full-length; full-length cDNA; Tetraodon nigroviridis.

REFERENCE
AUTHORS

1
Jailion,O., Aury,J.M., Brunet,F., Petit,J.L., Stange-Thomann,N.,
Mauceli,E., Bouneau,L., Fischer,C., Ozouf-Costaz,C., Bernot,A.,
Nicaud,S., Jaffe,D., Fisher,S., Lutfalla,G., Dossat,C.,
Segurens,B., Dasilva,C., Salanoubat,M., Levy,M., Boudet,N.,
Castellano,S., Anthouard,V., Jubin,C., Castellio,L., Poulain,J., De
Vacherie,B., Biemont,C., Skalli,Z., Cattolico,L., Coutanceau,J.P.,
Berardinis,V., Cruaud,C., Duprat,S., Brottier,P., Coutanceau,J.P.,
Gouzy,J., Parra,G., Lardier,G., Chapple,C., McKernan,K.J.,
McEwan,P., Bosak,S., Kellis,M., Volff,J.N., Guigo,R., Zody,M.C.,
Mesirov,J., Lindblad-Toh,K., Laudet,V., Schachter,V., Quetier,F.,
Robinson-Rechavi,M., Laudet,V., Schachter,V., Quetier,F.,
Saurin,W., ScarPELLI,C., Wincker,P., Lander,E.S., Weissbach,J.,
and Roest Crolius,H.
Genome duplication in the teleost fish Tetraodon nigroviridis
reveals the early vertebrate proto-karyotype
Nature 431 (7011), 946-957 (2004)
15496914
PUBMED
2 (bases 1 to 1454)
REFERENCE
AUTHORS

Direct Submission
Submitted (24-NOV-2004) Genoscope - Centre National de Sequencage -
: 2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
(E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
On Dec 3, 2004 this sequence version replaced gi:51194770.
The sequences are based on single pass reads.
More information available at
http://www.genoscope.cns.fr/tetraodon.
Location/Qualifiers
1. .1454
/organism="Tetraodon nigroviridis"
/mol_type="mRNA"
/db_xref="taxon:99883"
/tissue_type="Eyes"

ORIGIN

Query Match 90.6%; Score 15.4; DB 6; Length 1454;
Best Local Similarity 94.1%; Pred. No. 3.7e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCACGCTGGCCGAG 17
|||||||
Db 803 GCCACGCTAGCCGAG 819

RESULT 85
CNSOFYVC 1492 bp mRNA linear HTC 05-JUL-2005
LOCUS
DEFINITION Tetraodon nigroviridis full-length cDNA.
ACCESSION CR694556
VERSION CR694556.2 GI:56286437
KEYWORDS
SOURCE
ORGANISM

Htc; cDNA; full-length; full-length cDNA; Tetraodon nigroviridis.

REFERENCE
AUTHORS

1
Jailion,O., Aury,J.M., Brunet,F., Petit,J.L., Stange-Thomann,N.,
Mauceli,E., Bouneau,L., Fischer,C., Ozouf-Costaz,C., Bernot,A.,
Nicaud,S., Jaffe,D., Fisher,S., Lutfalla,G., Dossat,C.,
Segurens,B., Dasilva,C., Salanoubat,M., Levy,M., Boudet,N.,
Castellano,S., Anthouard,V., Jubin,C., Castellio,L., Poulain,J., De
Vacherie,B., Biemont,C., Skalli,Z., Cattolico,L., Coutanceau,J.P.,
Berardinis,V., Cruaud,C., Duprat,S., Brottier,P., Coutanceau,J.P.,
Gouzy,J., Parra,G., Lardier,G., Chapple,C., McKernan,K.J.,
McEwan,P., Bosak,S., Kellis,M., Volff,J.N., Guigo,R., Zody,M.C.,
Mesirov,J., Lindblad-Toh,K., Laudet,V., Schachter,V., Quetier,F.,
Robinson-Rechavi,M., Laudet,V., Schachter,V., Quetier,F.,
Saurin,W., ScarPELLI,C., Wincker,P., Lander,E.S., Weissbach,J.,
and Roest Crolius,H.
Genome duplication in the teleost fish Tetraodon nigroviridis
reveals the early vertebrate proto-karyotype
Nature 431 (7011), 946-957 (2004)
15496914
PUBMED
2 (bases 1 to 1492)
REFERENCE
AUTHORS

Direct Submission
Submitted (24-NOV-2004) Genoscope - Centre National de Sequencage -
: 2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE

```

COMMENT (E-mail : segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
On Dec 3, 2004 this sequence version replaced gi:51192463.  
The sequences are based on single pass reads.  
More information available at  
http://www.genoscope.cns.fr/tetraodon.

## FEATURES

source  
1.1492  
/organism="Tetraodon nigroviridis"  
/mol\_type="mRNA"  
/db\_xref="taxon:99883"  
/tissue\_type="Eyes"

## ORIGIN

Query Match 90.6%; Score 15.4; DB 6; Length 1492;  
Best Local Similarity 94.1%; Pred. No. 3.7e+03;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCCGAG 17  
|||||  
Db 818 GCCCAGCTAGCCGAG 834

## RESULT 86

CNS0GM2S 1505 bp mRNA linear HTC 05-JUL-2005  
LOCUS Tetraodon nigroviridis full-length cDNA.  
DEFINITION CR724638.2  
ACCESSION CR724638.2 GI:56306113  
VERSION HTG; cDNA; full-length; full-length cDNA; Tetraodon nigroviridis.  
KEYWORDS Tetraodon nigroviridis  
SOURCE Tetraodon nigroviridis  
ORGANISM

REFERENCE 1  
AUTHORS Jaillon,O., Aury,J.M., Brunet,F., Petit,J.L., Stange-Thomann,N., Mauceli,E., Bouneau,L., Fischer,C., Ozouf-Costaz,C., Bernot,A., Nicaud,S., Jaffe,D., Fisher,S., Lutfalla,G., Dossat,C., Segurens,B., Dasilva,C., Salanoubat,M., Levy,M., Boudet,N., Castellano,S., Anthonard,V., Jubin,C., Castelli,V., Katinka,M., Vacherie,B., Biemont,C., Skalli,Z., Cattolico,L., Poulain,J., De Berardinis,V., Cruaud,C., Duprat,S., Brottier,P., Coutanceau,J.P., Guzy,J., Parra,G., Lardier,G., Chapple,C., McKernan,K.J., McEwan,P., Bosak,S., Kellis,M., Volff,J.N., Guigo,R., Zody,M.C., Mesirov,J., Lindblad-Toh,K., Laudet,V., Schachter,V., Quetier,F., Robinson-Rechavi,M., Laudet,V., Schachter,V., Quetier,F., Saurin,W., Scarpelli,C., Wincker,P., Lander,E.S., Weissenbach,J. and Roest Crolius,H.  
TITLE Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype

JOURNAL Nature 431 (7011), 946-957 (2004)  
PUBMED 15496914  
REFERENCE 2 (bases 1 to 1505)  
AUTHORS Genoscope.  
TITLE Direct Submission  
JOURNAL Submitted (24-NOV-2004) Genoscope - Centre National de Sequenage - 2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE  
(E-mail : segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
On Dec 3, 2004 this sequence version replaced gi:51222889.  
COMMENT The sequences are based on single pass reads.  
More information available at  
http://www.genoscope.cns.fr/tetraodon.

FEATURES source  
1.1505  
/organism="Tetraodon nigroviridis"  
/mol\_type="mRNA"  
/db\_xref="taxon:99883"  
/tissue\_type="fish"

## ORIGIN

Query Match 90.6%; Score 15.4; DB 6; Length 1505;  
Best Local Similarity 94.1%; Pred. No. 3.7e+03;

Qy 1 GCCCAGCTTGGCCGAG 17  
|||||  
Db 837 GCCCAGCTAGCCGAG 853

## FEATURES

source  
1.1505  
/organism="Tetraodon nigroviridis"  
/mol\_type="mRNA"  
/db\_xref="taxon:99883"  
/tissue\_type="fish"

## ORIGIN

Query Match 90.6%; Score 15.4; DB 6; Length 1505;  
Best Local Similarity 94.1%; Pred. No. 3.7e+03;

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 GCCCAGCTTGGCCGAG 17  
|||||  
Db 822 GCCCAGCTAGCCGAG 838

## RESULT 87

CNS0G3DQ 1512 bp mRNA linear HTC 05-JUL-2005  
LOCUS Tetraodon nigroviridis full-length cDNA.  
DEFINITION CR700402  
ACCESSION CR700402  
VERSION HTG; cDNA; full-length; full-length cDNA; Tetraodon nigroviridis.  
KEYWORDS Tetraodon nigroviridis  
SOURCE Tetraodon nigroviridis  
ORGANISM

REFERENCE 1  
AUTHORS Jaillon,O., Aury,J.M., Brunet,F., Petit,J.L., Stange-Thomann,N., Mauceli,E., Bouneau,L., Fischer,C., Ozouf-Costaz,C., Bernot,A., Nicaud,S., Jaffe,D., Fisher,S., Lutfalla,G., Dossat,C., Segurens,B., Dasilva,C., Salanoubat,M., Levy,M., Boudet,N., Castellano,S., Anthonard,V., Jubin,C., Castelli,V., Katinka,M., Vacherie,B., Biemont,C., Skalli,Z., Cattolico,L., Poulain,J., De Berardinis,V., Cruaud,C., Duprat,S., Brottier,P., Coutanceau,J.P., Guzy,J., Parra,G., Lardier,G., Chapple,C., McKernan,K.J., McEwan,P., Bosak,S., Kellis,M., Volff,J.N., Guigo,R., Zody,M.C., Mesirov,J., Lindblad-Toh,K., Laudet,V., Schachter,V., Quetier,F., Robinson-Rechavi,M., Laudet,V., Schachter,V., Quetier,F., Saurin,W., Scarpelli,C., Wincker,P., Lander,E.S., Weissenbach,J. and Roest Crolius,H.  
TITLE Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype

JOURNAL Nature 431 (7011), 946-957 (2004)  
PUBMED 15496914  
REFERENCE 2 (bases 1 to 1512)  
AUTHORS Genoscope.  
TITLE Direct Submission  
JOURNAL Submitted (24-NOV-2004) Genoscope - Centre National de Sequenage - 2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE  
(E-mail : segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
On Dec 3, 2004 this sequence version replaced gi:51198311.  
COMMENT The sequences are based on single pass reads.  
More information available at  
http://www.genoscope.cns.fr/tetraodon.

## FEATURES

source  
1.1512  
/organism="Tetraodon nigroviridis"  
/mol\_type="mRNA"  
/db\_xref="taxon:99883"  
/tissue\_type="Eyes"

## ORIGIN

Query Match 90.6%; Score 15.4; DB 6; Length 1512;  
Best Local Similarity 94.1%; Pred. No. 3.7e+03;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCCGAG 17  
|||||  
Db 837 GCCCAGCTAGCCGAG 853

## RESULT 88

CNS0EXOO 1518 bp mRNA linear HTC 05-JUL-2005  
LOCUS Tetraodon nigroviridis full-length cDNA.  
DEFINITION CR646338  
ACCESSION CR646338  
VERSION HTG; cDNA; full-length; full-length cDNA; Tetraodon nigroviridis.  
KEYWORDS Tetraodon nigroviridis  
SOURCE

```

ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontoidea; Tetraodontidae; Tetraodon.

REFERENCE
AUTHORS Jaillon, O., Aury, J.M., Brunet, F., Petit, J.L., Stange-Thomann, N.,
Maucleri, E., Bouneau, L., Fischer, C., Ozouf-Costaz, C., Bernot, A.,
Nicaud, S., Jaffe, D., Fisher, S., Lutfalla, G., Dossat, C.,
Segurens, B., Da Silva, C., Salanoubat, M., Levy, M., Boudet, N.,
Castellano, S., Anthonard, V., Jubin, C., Castelli, V., Katinka, M.,
Vacherie, B., Blemont, C., Skalli, Z., Cattolico, L., Poulain, J., De
Berardinis, V., Cruaud, C., Duprat, S., Brottier, P., Coutanceau, J.P.,
Gouzy, J., Parra, G., Lardier, G., Chapple, C., McKernan, K.J.,
McEwan, P., Bosak, S., Kellis, M., Volff, J.N., Guigo, R., Zody, M.C.,
Mesirov, J., Lindblad-Toh, K., Birren, B., Nusbaum, C., Kahn, D.,
Robinson-Rechavi, M., Laudet, V., Schachter, V., Quetier, F.,
Saurin, W., Scarpelli, C., Wincker, P., Lander, E.S., Weissenbach, J.,
and Roest Crolius, H.
TITLE Genome duplication in the teleost fish Tetraodon nigroviridis
reveals the early vertebrate proto-karyotype
JOURNAL Nature 431 (7011), 946-957 (2004)
PUBMED 15496914
REFERENCE 2 (bases 1 to 1518)
Direct Submission
Genoscope.
Submitted (25-NOV-2004) Genoscope - Centre National de Sequencage -
: 2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
(E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
On Dec 3, 2004 this sequence version replaced gi:51142783.
The sequences are based on single pass reads.
More information available at
http://www.genoscope.cns.fr/tetraodon.

FEATURES
source
1..1518
/organism="Tetraodon nigroviridis"
/mol_type="mRNA"
/db_xref="taxon:99883"
/tissue_type="Liver"

ORIGIN
Query Match 90.6%; Score 15.4; DB 6; Length 1518;
Best Local Similarity 94.1%; Pred. No. 3.7e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCGCTTGGCCGAG 17
Db 841 GCCCAGCGCTAGGCCGAG 857

RESULT 89
DQ049847 1962 bp DNA linear GSS 02-JUN-2005
LOCUS DQ049847
DEFINITION Homo sapiens CLIP-2 gene, VIRTUAL TRANSCRIPT, partial sequence,
Genomic survey sequence.
ACCESSION DQ049847
VERSION DQ049847.1 GI:66903046
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS Hubisz, M.J., Flidel-Alon, A., Tanenbaum, D.M., Civello, D.,
Nielsen, R., Bustamante, C., Clark, A.G., Glanowski, S., Sackton, T.B.,
White, T.J., Sninsky, J.J., Adams, M.D. and Cargill, M.
TITLE A Scan for Positively Selected Genes in the Genomes of Humans and
Chimpanzees
JOURNAL PLoS Biol. 3 (6), E170 (2005)
PUBMED 15869325
REFERENCE 2 (bases 1 to 1962)
AUTHORS Nielsen, R., Bustamante, C., Clark, A.G., Glanowski, S., Sackton, T.B.,
White, T.J., Sninsky, J.J., Adams, M.D. and Cargill, M.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 4208)
Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raja, S.S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahney, J., Helton, E., Kettelman, M., Madan, A., Rodrigues, S.,
Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smal, D.E.,
Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Mammalian Gene Collection Program Team
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 4208)
.
NIH MGC Project
Direct Submission
Submitted (31-MAR-2004) National Institutes of Health, Mammalian
Gene Collection (MGC), Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk

```

```

Hubisz, M.J., Flidel-Alon, A., Tanenbaum, D.M., Civello, D.,
White, T.J., Sninsky, J.J., Adams, M.D. and Cargill, M.
Direct Submission
Submitted (05-MAY-2005) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
This sequence was made by sequencing genomic exons and ordering
them based on alignment. Translation starts at the beginning of
alignment.
FEATURES
source
1..1962
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="19"
<1..>1962
/genes="CLIP-2"
/locus_tag="HC17851"

ORIGIN
Query Match 90.6%; Score 15.4; DB 14; Length 1962;
Best Local Similarity 94.1%; Pred. No. 3.7e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCGCTTGGCCGAG 17
Db 1694 GCCCAGCGCTTGGCCAAG 1710

RESULT 90
BC068256 4208 bp mRNA linear HTC 12-DEC-2005
LOCUS BC068256
DEFINITION Homo sapiens cDNA clone IMAGE:5295699, containing frame-shift
errors.
ACCESSION BC068256
VERSION BC068256.1 GI:45946891
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 4208)
Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raja, S.S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahney, J., Helton, E., Kettelman, M., Madan, A., Rodrigues, S.,
Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smal, D.E.,
Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Mammalian Gene Collection Program Team
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 4208)
.
NIH MGC Project
Direct Submission
Submitted (31-MAR-2004) National Institutes of Health, Mammalian
Gene Collection (MGC), Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk

```

Email: cgapbs-re@mail.nih.gov  
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
 CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Toshiyuki and Piero Carninci (RIKEN)  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center  
 Center code: BCM-HGSC  
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
 Contact: amgobcm.tmc.edu  
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H., Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati, A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAC Plate: 91 Row: 1 Column: 14  
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 23510284  
 This clone has the following problem: frame shifted.

## FEATURES

Location/Qualifiers  
 1..4208  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5295699"  
 /tissue\_type="Testis"  
 /clone\_lib="NIH\_MGC\_97"  
 /lab\_host="DH10B"  
 /note="Vector: pbluescriptR"

## ORIGIN

Query Match 90.6%; Score 15.4; DB 6; Length 4208;  
 Best Local Similarity 94.1%; Pred. No. 3.7e+03;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCCGAG 17  
 |||||  
 Db 3283 GCCCAGCTTGGCCAG 3299

RESULT 91  
 B1529325  
 LOCUS  
 DEFINITION 1024097G09.yl C. reinhardtii CC-1690, normalized, Lambda Zap II  
 Chlamydomonas reinhardtii cDNA, mRNA sequence.

ACCESSION B1529325.1 GI:15369899  
 VERSION B1529325  
 KEYWORDS EST.

SOURCE  
 ORGANISM Chlamydomonas reinhardtii  
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;  
 Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.

REFERENCE 1 (bases 1 to 198)  
 Grossman, A., Chang, C.-W., Davies, J., Harris, E., Hauser, C., Lefebvre, P., McDermott, J. P., Shrager, J., Silflow, C. and Stern, D.  
 Analyses of the Chlamydomonas reinhardtii Genome: A Model, Unicellular System for Analyzing Gene Function and Regulation in Vascular Plants. Project: 1024b

JOURNAL  
 COMMENT Unpublished (2001)  
 Contact: Charles Hauser  
 DCMB Box 91000  
 Duke University  
 Durham, NC 27708-1000  
 Tel: 919 613 8159  
 Fax: 919 613 8177  
 Email: [chauser@duke.edu](mailto:chauser@duke.edu).

## FEATURES

Location/Qualifiers  
 1..198  
 /organism="Chlamydomonas reinhardtii"  
 /mol\_type="mRNA"  
 /strain="CC-1690 wild type mt+ 21gr"  
 /db\_xref="taxon:3055"

/clone\_lib="C. reinhardtii CC-1690, normalized, Lambda Zap II"  
 /note="vector: pbluescript II SK-; Site 1: EcoRI; Site 2: XhoI; This library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP (acetate-containing) medium in the light, TAP medium in the dark, HS (minimal) medium in ambient levels of CO2 and HS medium bubbled with 5% CO2. POLYA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. pBluescript II SK- plasmids were excised from the lambda Zap clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."

## ORIGIN

Query Match 88.2%; Score 15; DB 2; Length 198;  
 Best Local Similarity 100.0%; Pred. No. 5.8e+03;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 CCCACGCTTGCCGA 16  
 |||||  
 Db 100 CCCACGCTTGCCGA 114

## RESULT 92

A1613865/c  
 LOCUS  
 DEFINITION v988d01.yl Barstead mouse pooled organs MRLRB4 Mus musculus cDNA  
 clone IMAGE:873025 5', mRNA sequence.

ACCESSION A1613865.1 GI:4623032  
 VERSION A1613865  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 278)

Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.  
 The WashU-NCI Mouse EST Project 1999  
 Unpublished (1999)

CONTACT: Marra M/WashU-NCI Mouse EST Project 1999

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: [mousse@watson.wustl.edu](mailto:mousse@watson.wustl.edu)

This clone is available royalty-free through LLNL; contact the IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
 MGI:512505

This read is a RESEQUENCE of a previously sequenced mouse clone  
 This read has been verified (found to hit its original self in the correct orientation)

Seq primer: -40RP from Gibco

High quality sequence stop: 269.

## FEATURES

Location/Qualifiers  
 1..278  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:873025"  
 /sex="mixed"  
 /tissue\_type="pooled organs"  
 /dev\_stage="7 day"  
 /lab\_host="DH10B"  
 /clone\_lib="Barstead mouse pooled organs MRLRB4"

/note="Organ: pooled; Vector: pTT3D-PacI; Site 1: EcoRI; Site 2: NotI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTACGAATCTGAAGTGGGAGCGCGCCCTTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors [GTGATTCGGTACC], digested with Not I and cloned into the Not I and Eco RI sites of the modified pTT3 vector. Library constructed by Bob Barstead."

## ORIGIN

Query Match 88.2%; Score 15; DB 1; Length 278;  
Best Local Similarity 100.0%; Pred. No. 5.9e+03;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 CCACGCTTGGCCGAG 17  
|||||  
Db 250 CCACGCTTGGCCGAG 236

## RESULT 93

DR992377  
LOCUS M817011A24.g1 Apple\_EST\_Mdlr Malus x domestica cDNA similar to  
DEFINITION gb|AF68396.1| unknown protein [Oryza sativa (japonica  
cultivar-group)], mRNA sequence.

ACCESSION DR992377  
VERSION DR992377.1 GI:71814986  
KEYWORDS EST.  
SOURCE Malus x domestica

## ORGANISM

Malus x domestica  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
rosids; eurosids I; Rosales; Rosaceae; Maloideae; Malus.

## REFERENCE

1 (bases 1 to 309)  
Korban, S., Vodkin, L., Liu, L., Gasic, K., Gonzales, O., Hernandez, A.,  
Aldwinckle, H., Malnoy, M., Carroll, N., Goldsbrough, P., Orvis, K.,  
Clifton, S., Pape, D., Marra, M., Hillier, L., Martin, J., Wylie, T.,  
Dante, M., Theising, B., Bowers, Y., Gibbons, M., Ritter, E., Ronko, I.,  
Teagareihvili, R., Kennedy, S., Waterston, R., and Wilson, R.  
Apple Functional Genomics grant - NSF 0321702  
Unpublished (2004)

## AUTHORS

Contact: Schuyler S. Korban  
Apple Functional Genomics grant - NSF 0321702  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu

Library material provided by S. Korban/H. Aldwinckle/ M. Malnoy  
Library constructed by D.O.Gonzalez/L.Vodkin Library sequenced by  
Washington University Genome Sequencing Center  
Original QR value of 517 was extended to value 517 (,)

This trace has been recalled with phred

Original value before phred recall for SL was 112

Original value before phred recall for SR was 519.

## FEATURES

## source

1..309  
/organism="Malus x domestica"  
/mol\_type="mRNA"  
/cultivar="GoldRush"  
/db\_xref="taxon:3750"  
/tissue\_type="Leaves challenged with oblique banded leaf  
roller (Choristoneura rosaceana)"  
/lab\_host="DH10B ampicillin resistant"  
/clone\_lib="Apple\_EST\_Mdlr"  
/note="Vector: pBluescript II SK (+); Site 1: XhoI;  
Site 2: EcoRI; Total RNA was extracted from freeze dried  
leaf tissue, using the method described by Wang and Vodkin  
(Plant Molecular Biology Reporter 12:132-145, 1994).  
Poly(A) + mRNA was isolated from total RNA using the  
PolyATract mRNA Isolation system III (Promega). The  
library was prepared using the Stratagene pBluescript II  
XR cDNA library construction kit. Complementary DNA was

## ORIGIN

Query Match 88.2%; Score 15; DB 10; Length 309;  
Best Local Similarity 100.0%; Pred. No. 5.9e+03;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 CCACGCTTGGCCGAG 17  
|||||  
Db 203 CCACGCTTGGCCGAG 217

## RESULT 94

AV850077  
LOCUS AV850077  
DEFINITION AV850077 Nori Satoh unpublished cDNA library, cleavage stage embryo  
Ciona intestinalis cDNA clone rcic104n18 3', mRNA sequence.

ACCESSION AV850077  
VERSION AV850077.1 GI:16832959  
KEYWORDS EST.  
SOURCE Ciona intestinalis

## ORGANISM

Ciona intestinalis  
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;  
Phlebobranchia; Cloniidae; Ciona.

## REFERENCE

1 (bases 1 to 322)  
Satoh, N., Satou, Y., Kohara, Y. and Shin-i, T.

Expressed genes in Ciona intestinalis

Unpublished (2000)

Contact: Nori Satoh

Department of Zoology

Kyoto University

Sakyo-ku, Kyoto, Kyoto 606-8502, Japan

Tel: 81-75-753-4081

Fax: 81-75-705-1113

Email: satoh@ascidian.zool.kyoto-u.ac.jp.

## FEATURES

## source

1..322  
/organism="Ciona intestinalis"  
/mol\_type="mRNA"  
/db\_xref="taxon:7719"  
/clone="rcic104n18"  
/tissue\_type="whole animal"  
/dev\_stage="cleaving embryo"  
/clone\_lib="Nori Satoh unpublished cDNA library, cleavage  
stage embryo"

## ORIGIN

Query Match 88.2%; Score 15; DB 7; Length 322;  
Best Local Similarity 100.0%; Pred. No. 5.9e+03;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 CCCACGCTTGGCCGA 16  
|||||  
Db 146 CCCACGCTTGGCCGA 160

## RESULT 95

CA661056  
LOCUS wlm1.pk0017.f2 wlm1 Triticum aestivum cDNA clone wlm1.pk0017.f2 5'  
DEFINITION end, mRNA sequence.

ACCESSION CA661056

VERSION CA661056.1 GI:25239581

KEYWORDS EST.

synthesized from mRNA using a Poly (dT) sequence with a  
XhoI restriction site. EcoRI adapters were ligated to the  
blunt ended cDNA fragments followed by restriction with  
XhoI. The cDNA insert is protected from XhoI digestion via  
methylation during first strand cDNA synthesis. The cDNA  
fragments were directionally cloned into the EcoRI-XhoI  
restriction site of the pBluescript vector. The ligated  
cDNA fragments were transformed into E. coli Electromax  
DH10B host cells. Transformation efficiency: 4.0E+04  
Average insert Size by PCR: 1320 bp"

```

SOURCE
ORGANISM      Triticum aestivum (bread wheat)
               Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
               Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP
               clade; Pooideae; Triticeae; Triticum.
REFERENCE
AUTHORS       1 (bases 1 to 368)
               Tingey,S.V., Powell,W., Wolters,P., Dolan,M., Hainey,C., Yuan,Z.,
               Miao,G., Caraher,N. and Hanafey,M.K.
TITLE         DuPont Wheat cDNA Sequence
JOURNAL       Unpublished (2002)
COMMENT       Contact: Scott V. Tingey
               Crop Genetics
               E. I. DuPont de Nemours and Company
               1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA
               Tel: 302-631-2602
               Fax: 302-631-2607
               Email: Scott.V.tingey@USA.dupont.com
               Seq primer: M13.

FEATURES
source
1..368
   /organism="Triticum aestivum"
   /mol_type="mRNA"
   /cultivar="Stephens"
   /db_xref="taxon:4565"
   /clone="wlm1.pk0017.f2"
   /tissue_type="leaf"
   /clone_lib="wlm1"
   /note="Vector: pBluescript SK+; Site 1: EcoRI; Site 2:
   XhoI; Wheat (Triticum aestivum L.) seedlings 1 hr after
   inoculation with Erysiphe graminis f. sp tritici"

ORIGIN
Query Match      88.2%; Score 15; DB 4; Length 368;
Best Local Similarity 100.0%; Pred. No. 5.9e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 CCACGCTTGGCCGAG 17
   |||||
Db 295 CCACGCTTGGCCGAG 309

RESULT 96
BW421246
LOCUS
DEFINITION      BW421246 Yutaka Satou unpublished cDNA library, mature adult whole
animal Ciona intestinalis CDNA clone cima839j04 3', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM      Ciona intestinalis
               Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
               Phlebobranchia; Cionidae; Ciona.
REFERENCE
AUTHORS       1 (bases 1 to 419)
               Satou,Y., Shin-I,T., Kohara,Y. and Satoh,N.
TITLE         Expressed genes in Ciona intestinalis (2004)
JOURNAL       Unpublished (2004)
COMMENT       Contact: Yutaka Satou
               Department of Zoology
               Kyoto University
               Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
               Tel: 81-75-753-4095
               Fax: 81-75-705-1113
               Email: yutaka@aeicdian.zool.kyoto-u.ac.jp.
               Location/Qualifiers
               1..419
               /organism="Ciona intestinalis"
               /mol_type="mRNA"
               /db_xref="taxon:7719"
               /clone="cima839j04"
               /tissue_type="whole animal"
               /dev_stage="mature adult"
               /clone_lib="Yutaka Satou unpublished cDNA library, mature

SOURCE
ORGANISM      Triticum aestivum (bread wheat)
               Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
               Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP
               clade; Pooideae; Triticeae; Triticum.
REFERENCE
AUTHORS       1 (bases 1 to 368)
               Tingey,S.V., Powell,W., Wolters,P., Dolan,M., Hainey,C., Yuan,Z.,
               Miao,G., Caraher,N. and Hanafey,M.K.
TITLE         DuPont Wheat cDNA Sequence
JOURNAL       Unpublished (2002)
COMMENT       Contact: Scott V. Tingey
               Crop Genetics
               E. I. DuPont de Nemours and Company
               1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA
               Tel: 302-631-2602
               Fax: 302-631-2607
               Email: Scott.V.tingey@USA.dupont.com
               Seq primer: M13.

FEATURES
source
1..368
   /organism="Triticum aestivum"
   /mol_type="mRNA"
   /cultivar="Stephens"
   /db_xref="taxon:4565"
   /clone="wlm1.pk0017.f2"
   /tissue_type="leaf"
   /clone_lib="wlm1"
   /note="Vector: pBluescript SK+; Site 1: EcoRI; Site 2:
   XhoI; Wheat (Triticum aestivum L.) seedlings 1 hr after
   inoculation with Erysiphe graminis f. sp tritici"

ORIGIN
Query Match      88.2%; Score 15; DB 4; Length 368;
Best Local Similarity 100.0%; Pred. No. 5.9e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 CCACGCTTGGCCGAG 17
   |||||
Db 295 CCACGCTTGGCCGAG 309

RESULT 96
BW421246
LOCUS
DEFINITION      BW421246 Yutaka Satou unpublished cDNA library, mature adult whole
animal Ciona intestinalis CDNA clone cima839j04 3', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM      Ciona intestinalis
               Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
               Phlebobranchia; Cionidae; Ciona.
REFERENCE
AUTHORS       1 (bases 1 to 419)
               Satou,Y., Shin-I,T., Kohara,Y. and Satoh,N.
TITLE         Expressed genes in Ciona intestinalis (2004)
JOURNAL       Unpublished (2004)
COMMENT       Contact: Yutaka Satou
               Department of Zoology
               Kyoto University
               Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
               Tel: 81-75-753-4095
               Fax: 81-75-705-1113
               Email: yutaka@aeicdian.zool.kyoto-u.ac.jp.
               Location/Qualifiers
               1..419
               /organism="Ciona intestinalis"
               /mol_type="mRNA"
               /db_xref="taxon:7719"
               /clone="cima839j04"
               /tissue_type="whole animal"
               /dev_stage="mature adult"
               /clone_lib="Yutaka Satou unpublished cDNA library, mature

```

```

ORIGIN
adult whole animal"

Query Match      88.2%; Score 15; DB 3; Length 419;
Best Local Similarity 100.0%; Pred. No. 5.9e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 CCCACGCTTGGCCGAG 16
   |||||
Db 188 CCCACGCTTGGCCGAG 202

RESULT 97
CN899452
LOCUS
DEFINITION      CN899452 438 bp mRNA linear EST 04-JUN-2004
010706ABAA999039HT (ABAA) Pacific Rose spur buds from ON trees
Malus x domestica cDNA clone ABAA999039, mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM      Malus x domestica
               Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
               Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
               rosids; eurosids I; Rosales; Rosaceae; Maloideae; Malus.
REFERENCE
AUTHORS       1 (bases 1 to 438)
               Beuning,L., Bowen,J., Crowhurst,R., Gleave,A., Janssen,B.,
               McArtney,S., Newcomb,R., Ross,G., Snowden,K., Walton,E. and Yauk,Y.
               HortResearch Apple EST Project
               Unpublished (2004)
               Contact: Gleave,A.
               Sequencing Facility
               The Horticulture and Food Research Institute of New Zealand Ltd
               120 Mc Albert Rd, Mt Albert, Auckland, New Zealand
               Tel: 00 64 09 815 4200
               Fax: 00 64 09 815 4201
               Email: est@hortresearch.co.nz.
               Location/Qualifiers
               1..438
               /organism="Malus x domestica"
               /mol_type="mRNA"
               /db_xref="taxon:3750"
               /clone="ABAA999039"
               /tissue_type="Spur buds from 'ON' trees"
               /clone_lib="(ABAA) Pacific Rose spur buds from ON trees"
               /note="Vector: pBluescript SK(-); Library sequenced by
               Genesis Research & Development"

ORIGIN
Query Match      88.2%; Score 15; DB 8; Length 438;
Best Local Similarity 100.0%; Pred. No. 5.9e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 CCACGCTTGGCCGAG 17
   |||||
Db 175 CCACGCTTGGCCGAG 189

RESULT 98
BW366787/c
LOCUS
DEFINITION      BW366787 Yutaka Satou unpublished cDNA library, mature adult whole
animal Ciona intestinalis cDNA clone cima839j04 5', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM      Ciona intestinalis
               Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
               Phlebobranchia; Cionidae; Ciona.
REFERENCE
AUTHORS       1 (bases 1 to 484)
               Satou,Y., Shin-I,T., Kohara,Y. and Satoh,N.
TITLE         Expressed genes in Ciona intestinalis (2004)

```

```
JOURNAL Unpublished (2004)
COMMENT Contact: Yutaka Satou
          Department of Zoology
          Kyoto University
          Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
          Tel: 81-75-753-4095
          Fax: 81-75-705-1113
          Email: yutaka@ascidian.zool.kyoto-u.ac.jp.

FEATURES
source 1..484
        /organism="Ciona intestinalis"
        /mol_type="mRNA"
        /db_xref="taxon:7719"
        /clone="cima839j04"
        /tissue_type="whole animal"
        /dev_stage="mature adult"
        /clone_lib="Yutaka Satou unpublished cDNA library, mature
        adult whole animal"

ORIGIN
Query Match 88.2%; Score 15; DB 3; Length 484;
Best Local Similarity 100.0%; Pred. No. 5.9e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 CCCACGCTTGCGCGA 16
Db 232 CCCACGCTTGCGCGA 218
|||||

RESULT 99
BW178352 492 bp mRNA linear EST 04-NOV-2002
LOCUS BW178352 Nori Satoh unpublished cDNA library, heart Ciona
DEFINITION intestinalis cDNA clone rcit005e16 3', mRNA sequence.
ACCESSION BW178352
VERSION BW178352.1 GI:24568276
KEYWORDS EST.
SOURCE Ciona intestinalis
ORGANISM Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
          Phlebobranchia; Cionidae; Ciona.
REFERENCE 1 (bases 1 to 492)
AUTHORS Satou,Y., Shin-i,T., Kohara,Y. and Satoh,N.
TITLE Expressed genes in Ciona intestinalis (2002c)
JOURNAL Unpublished (2002)
COMMENT Contact: Nori Satoh
          Department of Zoology
          Kyoto University
          Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
          Tel: 81-75-753-4081
          Fax: 81-75-705-1113
          Email: satoh@ascidian.zool.kyoto-u.ac.jp.

FEATURES
source 1..492
        /organism="Ciona intestinalis"
        /mol_type="mRNA"
        /db_xref="taxon:7719"
        /clone="rcit005e16"
        /tissue_type="heart"
        /clone_lib="Nori Satoh unpublished cDNA library, heart"

ORIGIN
Query Match 88.2%; Score 15; DB 3; Length 492;
Best Local Similarity 100.0%; Pred. No. 5.9e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 CCCACGCTTGCGCGA 16
Db 227 CCCACGCTTGCGCGA 241
|||||

RESULT 100
BW203028/c
```

---

```
LOCUS BW203028 517 bp mRNA linear EST 01-JUN-2005
DEFINITION BW203028 Nori Satoh unpublished cDNA library, cleaving embryo Ciona
            intestinalis cDNA clone cic1083c24 5', mRNA sequence.
ACCESSION BW203028
VERSION BW203028.1 GI:24617457
KEYWORDS EST.
SOURCE Ciona intestinalis
ORGANISM Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
          Phlebobranchia; Cionidae; Ciona.
REFERENCE 1 (bases 1 to 517)
AUTHORS Satou,Y., Shin-i,T., Kohara,Y. and Satoh,N.
TITLE Expressed genes in Ciona intestinalis (2002c)
JOURNAL Unpublished (2002)
COMMENT Contact: Nori Satoh
          Department of Zoology
          Kyoto University
          Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
          Tel: 81-75-753-4081
          Fax: 81-75-705-1113
          Email: satoh@ascidian.zool.kyoto-u.ac.jp.

FEATURES
source 1..517
        /organism="Ciona intestinalis"
        /mol_type="mRNA"
        /db_xref="taxon:7719"
        /clone="cic1083c24"
        /tissue_type="whole body"
        /dev_stage="cleaving embryo"
        /clone_lib="Nori Satoh unpublished cDNA library, cleaving
        embryo"

ORIGIN
Query Match 88.2%; Score 15; DB 3; Length 517;
Best Local Similarity 100.0%; Pred. No. 5.9e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 CCCACGCTTGCGCGA 16
Db 275 CCCACGCTTGCGCGA 261
|||||

Search completed: June 10, 2006, 19:37:09
Job time : 2150.8 secs
```



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2006, 15:21:38 ; Search time 38,1083 Seconds  
(without alignments)  
834.695 Million cell updates/sec

Title: US-10-600-816-31

Perfect score: 17

Sequence: 1 gccacgccttgccgag 17

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1403666 seqs, 935554401 residues

Total number of hits satisfying chosen parameters: 2807332

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

Database :

1: /EMC\_Celerra\_SIDS3/ptodata/2/ina/1 COMB.seq.\*

2: /EMC\_Celerra\_SIDS3/ptodata/2/ina/5 COMB.seq.\*

3: /EMC\_Celerra\_SIDS3/ptodata/2/ina/6A COMB.seq.\*

4: /EMC\_Celerra\_SIDS3/ptodata/2/ina/6B COMB.seq.\*

5: /EMC\_Celerra\_SIDS3/ptodata/2/ina/7 COMB.seq.\*

6: /EMC\_Celerra\_SIDS3/ptodata/2/ina/H COMB.seq.\*

7: /EMC\_Celerra\_SIDS3/ptodata/2/ina/PCTUS COMB.seq.\*

8: /EMC\_Celerra\_SIDS3/ptodata/2/ina/PP COMB.seq.\*

9: /EMC\_Celerra\_SIDS3/ptodata/2/ina/RE COMB.seq.\*

10: /EMC\_Celerra\_SIDS3/ptodata/2/ina/backfileseq1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	17	100.0	601	3	US-09-949-016-29173
2	17	100.0	601	3	US-09-949-016-126957
3	17	100.0	1212	3	US-09-188-930-249
4	17	100.0	1212	3	US-09-312-283C-249
5	17	100.0	1228	3	US-09-495-050A-223
6	17	100.0	2289	3	US-09-949-016-3613
7	17	100.0	2456	3	US-09-949-016-572
8	17	100.0	26086	3	US-09-949-016-15355
9	17	100.0	26238	3	US-09-949-016-12314
10	15.4	90.6	1527	3	US-09-252-991A-11321
11	15.4	90.6	4183	3	US-08-996-083-2
12	15.4	90.6	4183	3	US-09-429-016-2
13	14.4	84.7	601	3	US-09-949-016-157184
14	14.4	84.7	1020	3	US-09-027-900-9
15	14.4	84.7	1520	3	US-09-027-900-8
16	14.4	84.7	1675	3	US-09-027-900-7
17	14.4	84.7	42242	3	US-09-949-016-16127
18	14.4	84.7	47698	3	US-09-949-016-13514
19	14.4	84.7	49312	3	US-09-671-317-485
20	14.4	84.7	67002	3	US-09-949-016-16803
21	13.8	81.2	20	3	US-09-332-522E-17
22	13.8	81.2	47	3	US-09-422-978-2426
23	13.8	81.2	283	3	US-09-313-294A-2571

24	13.8	81.2	293	3	US-09-313-294A-5109	Sequence 5109, Ap
25	13.8	81.2	385	3	US-08-990-823-44	Sequence 44, Appl
26	13.8	81.2	385	3	US-09-477-135A-44	Sequence 44, Appl
27	13.8	81.2	600	4	US-10-233-942-3	Sequence 3, Appl1
28	13.8	81.2	600	5	US-09-945-952A-3	Sequence 3, Appl1
29	13.8	81.2	601	3	US-09-949-016-121284	Sequence 121284,
30	13.8	81.2	601	3	US-09-949-016-205926	Sequence 205926,
31	13.8	81.2	601	3	US-09-949-016-205927	Sequence 205927,
32	13.8	81.2	795	3	US-09-902-540-5464	Sequence 5464, Ap
33	13.8	81.2	999	3	US-09-252-991A-6261	Sequence 6261, Ap
34	13.8	81.2	1176	3	US-09-252-991A-547	Sequence 547, App
35	13.8	81.2	1182	3	US-09-540-236-266	Sequence 266, App
36	13.8	81.2	1235	2	US-08-095-726-13	Sequence 13, Appl
37	13.8	81.2	1235	2	US-08-095-726-15	Sequence 15, Appl
38	13.8	81.2	1235	2	US-08-096-623A-13	Sequence 13, Appl
39	13.8	81.2	1235	2	US-08-096-623A-15	Sequence 15, Appl
40	13.8	81.2	1251	3	US-09-252-991A-1019	Sequence 1019, Ap
41	13.8	81.2	1578	2	US-08-681-129-1	Sequence 1, Appl1
42	13.8	81.2	1608	3	US-09-252-991A-569	Sequence 569, App
43	13.8	81.2	1616	3	US-09-869-588-32	Sequence 32, Appl
44	13.8	81.2	1653	3	US-09-252-991A-1111	Sequence 1111, Ap
45	13.8	81.2	1976	3	US-09-165-042-2	Sequence 2, Appl1
46	13.8	81.2	1983	3	US-09-252-991A-531	Sequence 531, App
47	13.8	81.2	2049	3	US-09-252-991A-6426	Sequence 6426, Ap
48	13.8	81.2	2058	2	US-08-358-117-1	Sequence 1, Appl1
49	13.8	81.2	2058	3	US-08-470-588-1	Sequence 1, Appl1
50	13.8	81.2	2169	3	US-09-976-594-315	Sequence 315, App
51	13.8	81.2	2174	3	US-09-613-444-1	Sequence 1, Appl1
52	13.8	81.2	2174	3	US-10-142-516-1	Sequence 1, Appl1
53	13.8	81.2	2184	3	US-09-902-540-5116	Sequence 5116, Ap
54	13.8	81.2	2847	3	US-09-252-991A-1036	Sequence 1036, Ap
55	13.8	81.2	3133	3	US-08-162-809-1	Sequence 1, Appl1
56	13.8	81.2	3453	3	US-09-902-540-4713	Sequence 4713, Ap
57	13.8	81.2	3636	4	US-10-094-039-957	Sequence 957, App
58	13.8	81.2	3814	3	US-09-919-079-254	Sequence 254, App
59	13.8	81.2	4563	3	US-09-252-991A-930	Sequence 930, App
60	13.8	81.2	5382	3	US-09-479-122-21	Sequence 21, Appl
61	13.8	81.2	5382	3	US-09-484-997-21	Sequence 21, Appl
62	13.8	81.2	5382	3	US-09-481-355-21	Sequence 21, Appl
63	13.8	81.2	5382	3	US-09-481-282-21	Sequence 21, Appl
64	13.8	81.2	5382	3	US-09-455-659A-21	Sequence 21, Appl
65	13.8	81.2	5382	3	US-09-484-996-21	Sequence 21, Appl
66	13.8	81.2	5382	3	US-09-479-123-21	Sequence 21, Appl
67	13.8	81.2	5382	3	US-09-484-317A-21	Sequence 21, Appl
68	13.8	81.2	5382	3	US-09-276-820A-21	Sequence 21, Appl
69	13.8	81.2	5382	5	US-09-760-897A-21	Sequence 21, Appl
70	13.8	81.2	8540	3	US-08-487-283A-4	Sequence 4, Appl1
71	13.8	81.2	8540	7	PCT-US96-05611A-12	Sequence 12, Appl
72	13.8	81.2	9737	3	US-09-479-122-22	Sequence 22, Appl
73	13.8	81.2	9737	3	US-09-479-122-23	Sequence 23, Appl
74	13.8	81.2	9737	3	US-09-484-997-22	Sequence 22, Appl
75	13.8	81.2	9737	3	US-09-484-997-23	Sequence 22, Appl
76	13.8	81.2	9737	3	US-09-484-997-28	Sequence 28, Appl
77	13.8	81.2	9737	3	US-09-484-997-28	Sequence 28, Appl
78	13.8	81.2	9737	3	US-09-481-355-22	Sequence 22, Appl
79	13.8	81.2	9737	3	US-09-481-355-23	Sequence 23, Appl
80	13.8	81.2	9737	3	US-09-481-355-28	Sequence 28, Appl
81	13.8	81.2	9737	3	US-09-481-282-22	Sequence 22, Appl
82	13.8	81.2	9737	3	US-09-481-282-23	Sequence 23, Appl
83	13.8	81.2	9737	3	US-09-481-282-28	Sequence 28, Appl
84	13.8	81.2	9737	3	US-09-455-659A-22	Sequence 22, Appl
85	13.8	81.2	9737	3	US-09-455-659A-23	Sequence 23, Appl
86	13.8	81.2	9737	3	US-09-455-659A-28	Sequence 28, Appl
87	13.8	81.2	9737	3	US-09-484-996-22	Sequence 22, Appl
88	13.8	81.2	9737	3	US-09-484-996-23	Sequence 23, Appl
89	13.8	81.2	9737	3	US-09-484-996-28	Sequence 28, Appl
90	13.8	81.2	9737	3	US-09-479-123-22	Sequence 22, Appl
91	13.8	81.2	9737	3	US-09-479-123-23	Sequence 23, Appl
92	13.8	81.2	9737	3	US-09-479-123-28	Sequence 28, Appl
93	13.8	81.2	9737	3	US-09-484-317A-22	Sequence 22, Appl
94	13.8	81.2	9737	3	US-09-484-317A-23	Sequence 23, Appl
95	13.8	81.2	9737	3	US-09-484-317A-28	Sequence 28, Appl
96	13.8	81.2	9737	3	US-09-276-820A-22	Sequence 22, Appl















c 973 12.4 72.9 203475 3 US-09-949-016-17229  
c 974 12.4 72.9 205044 3 US-09-949-016-15851  
c 975 12.4 72.9 205044 3 US-09-949-016-15852  
c 976 12.4 72.9 205044 3 US-09-949-016-15853  
c 977 12.4 72.9 219964 3 US-09-949-016-15086  
c 978 12.4 72.9 223471 3 US-09-949-016-12387  
c 979 12.4 72.9 223471 3 US-09-949-016-12724  
c 980 12.4 72.9 223471 3 US-09-949-016-12725  
c 981 12.4 72.9 227979 3 US-09-949-016-11842  
c 982 12.4 72.9 227979 3 US-09-949-016-12706  
c 983 12.4 72.9 275110 3 US-09-949-016-16070  
c 984 12.4 72.9 283538 3 US-09-949-016-13506  
c 985 12.4 72.9 300598 3 US-09-949-016-11868  
c 986 12.4 72.9 302604 3 US-09-949-016-14589  
c 987 12.4 72.9 302604 3 US-09-949-016-14589  
c 988 12.4 72.9 304533 3 US-09-949-016-15371  
c 989 12.4 72.9 304533 3 US-09-949-016-15372  
c 990 12.4 72.9 308362 3 US-09-949-016-17119  
c 991 12.4 72.9 319608 3 US-09-539-333D-1  
c 992 12.4 72.9 319608 3 US-09-679-409-1  
c 993 12.4 72.9 374159 3 US-09-949-016-15868  
c 994 12.4 72.9 450395 3 US-09-949-016-15473  
c 995 12.4 72.9 536165 3 US-09-214-808-1  
c 996 12.2 71.8 17 3 US-09-866-108A-9670  
c 997 12.2 71.8 25 3 US-09-866-108A-14562  
c 998 12.2 71.8 25 3 US-09-866-108A-14563  
c 999 12.2 71.8 25 3 US-09-866-108A-14564  
1000 12.2 71.8 25 3 US-09-866-108A-14565

## ALIGNMENTS

RESULT 1  
US-09-949-016-29173  
; Sequence 29173, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 29173  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-29173

Query Match 100.0%; Score 17; DB 3; Length 601;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17  
Db 14 GCCCAGCGTTGGCCGAG 30

RESULT 2  
US-09-949-016-126957  
; Sequence 126957, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 126957  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-126957

Query Match 100.0%; Score 17; DB 3; Length 601;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17  
Db 14 GCCCAGCGTTGGCCGAG 30

RESULT 3  
US-09-188-930-249  
; Sequence 249, Application US/09188930A  
; Patent No. 6150502  
; GENERAL INFORMATION:  
; APPLICANT: Watson, James D.  
; APPLICANT: Strachan, Lorna  
; APPLICANT: Sleeman, Matthew  
; APPLICANT: Orrust, Rene  
; APPLICANT: Murison, James Greg  
; TITLE OF INVENTION: Compositions Isolated From Skin Cells  
; TITLE OF INVENTION: and Methods For Their Use  
; FILE REFERENCE: 11000.1011c1  
; CURRENT APPLICATION NUMBER: US/09/188,930A  
; CURRENT FILING DATE: 1998-11-09  
; NUMBER OF SEQ ID NOS: 348  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 249  
; LENGTH: 1212  
; TYPE: DNA  
; ORGANISM: Human  
US-09-188-930-249

Query Match 100.0%; Score 17; DB 3; Length 1212;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17  
Db 990 GCCCAGCGTTGGCCGAG 1006

RESULT 4  
US-09-312-283C-249  
; Sequence 249, Application US/09312283C  
; Patent No. 6573095  
; GENERAL INFORMATION:  
; APPLICANT: Watson, James D.  
; APPLICANT: Strachan, Lorna  
; APPLICANT: Sleeman, Matthew  
; APPLICANT: Orrust, Rene  
; APPLICANT: Murison, James G.  
; APPLICANT: Kumble, Krishanand D.  
; TITLE OF INVENTION: Compositions Isolated from Skin Cells  
; TITLE OF INVENTION: and Methods for Their Use



; FILE REFERENCE: 11000.1011c2  
; CURRENT APPLICATION NUMBER: US/09/312,283C  
; CURRENT FILING DATE: 1999-05-14  
; NUMBER OF SEQ ID NOS: 425  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 249  
; LENGTH: 1212  
; TYPE: DNA  
; ORGANISM: Mouse  
US-09-312-283C-249

Query Match 100.0%; Score 17; DB 3; Length 1212;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCCAGCTTGCCGCGAG 17  
|||  
DB 990 GCCCAGCTTGCCGCGAG 1006

## RESULT 5

US-09-495-050A-223  
; Sequence 223, Application US/09495050A  
; Patent No. 6492505

; GENERAL INFORMATION:  
; APPLICANT: Roopa, Reddy

; APPLICANT: Guegler, Karl, J.  
; APPLICANT: Au-Young, Janice

; TITLE OF INVENTION: COMPOSITION FOR DETECTION OF GENES ENCODING MEMBRANE-ASSOCIATED P

; FILE REFERENCE: PA-0013 US  
; CURRENT APPLICATION NUMBER: US/09/495,050A

; CURRENT FILING DATE: 2000-01-31

; PRIOR APPLICATION NUMBER: 60/118,318

; PRIOR FILING DATE: February 1, 1999

; NUMBER OF SEQ ID NOS: 305

; SOFTWARE: PERL Program

; SEQ ID NO 223

; LENGTH: 1228

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc.feature

; OTHER INFORMATION: Incyte ID No. 6492505 2681738CT1

US-09-495-050A-223

Query Match 100.0%; Score 17; DB 3; Length 1228;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCCAGCTTGCCGCGAG 17  
|||  
DB 541 GCCCAGCTTGCCGCGAG 557

## RESULT 6

US-09-949-016-15355  
; Sequence 3613, Application US/09949016  
; Patent No. 6812339

; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

; FILE REFERENCE: CL001307

; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 2000-04-14

; PRIOR APPLICATION NUMBER: 60/241,755

; PRIOR FILING DATE: 2000-10-20

; PRIOR APPLICATION NUMBER: 60/237,768

; PRIOR FILING DATE: 2000-10-03

; PRIOR APPLICATION NUMBER: 60/231,498

; PRIOR FILING DATE: 2000-09-08

; NUMBER OF SEQ ID NOS: 207012

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 3613  
; LENGTH: 2289  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-3613

Query Match 100.0%; Score 17; DB 3; Length 2289;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCCAGCTTGCCGCGAG 17  
|||  
DB 1117 GCCCAGCTTGCCGCGAG 1133

## RESULT 7

US-09-949-016-572

; Sequence 572, Application US/09949016  
; Patent No. 6812339

; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

; FILE REFERENCE: CL001307

; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 2000-04-14

; PRIOR APPLICATION NUMBER: 60/241,755

; PRIOR FILING DATE: 2000-10-20

; PRIOR APPLICATION NUMBER: 60/237,768

; PRIOR FILING DATE: 2000-10-03

; PRIOR APPLICATION NUMBER: 60/231,498

; PRIOR FILING DATE: 2000-09-08

; NUMBER OF SEQ ID NOS: 207012

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 572

; LENGTH: 2456

; TYPE: DNA

; ORGANISM: Human

US-09-949-016-572

Query Match 100.0%; Score 17; DB 3; Length 2456;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCCAGCTTGCCGCGAG 17  
|||  
DB 1271 GCCCAGCTTGCCGCGAG 1287

## RESULT 8

US-09-949-016-15355

; Sequence 15355, Application US/09949016  
; Patent No. 6812339

; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

; FILE REFERENCE: CL001307

; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 2000-04-14

; PRIOR APPLICATION NUMBER: 60/241,755

; PRIOR FILING DATE: 2000-10-20

; PRIOR APPLICATION NUMBER: 60/237,768

; PRIOR FILING DATE: 2000-10-03

; PRIOR APPLICATION NUMBER: 60/231,498

; PRIOR FILING DATE: 2000-09-08

; NUMBER OF SEQ ID NOS: 207012

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 15355

; LENGTH: 26086

; TYPE: DNA

; ORGANISM: Human

US-09-949-016-15355

```
Query Match      100.0%; Score 17; DB 3; Length 26086;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCACGCTTGCCGAG 17
    |||||
Db 22913 GCCCACGCTTGCCGAG 22929

RESULT 9
US-09-949-016-12314
; Sequence 12314, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12314
; LENGTH: 26238
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12314

Query Match      100.0%; Score 17; DB 3; Length 26238;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCACGCTTGCCGAG 17
    |||||
Db 23067 GCCCACGCTTGCCGAG 23083

RESULT 10
US-09-252-991A-11321
; Sequence 11321, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11321
; LENGTH: 1527
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11321

Query Match      90.6%; Score 15.4; DB 3; Length 1527;
Best Local Similarity 94.1%; Pred. No. 1.3e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCACGCTTGCCGAG 17
    |||||
Db 25 GCCCACGCTTGCCGAG 41
```

```
RESULT 11
US-08-996-083-2
; Sequence 2, Application US/08996083A
; Patent No. 6124095
; GENERAL INFORMATION:
; APPLICANT: Magna, Holly
; APPLICANT: Schaffer, Paul
; APPLICANT: Lawton, Michael
; APPLICANT: Yocum, Sue
; APPLICANT: Mitchell, Peter
; APPLICANT: Hutchinson, Nancy
; APPLICANT: Murry, Lynn E.
; TITLE OF INVENTION: HUMAN NUCLEOTIDE PYROPHOSPHOHYDROLASE-2
; FILE REFERENCE: PF-0420 US
; CURRENT APPLICATION NUMBER: US/08/996,083A
; CURRENT FILING DATE: 1997-12-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0.
; SEQ ID NO 2
; LENGTH: 4183
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte Clone No. 6124095: 1388013
; PUBLICATION INFORMATION:
US-08-996-083-2

Query Match      90.6%; Score 15.4; DB 3; Length 4183;
Best Local Similarity 94.1%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCACGCTTGCCGAG 17
    |||||
Db 3254 GCCCACGCTTGCCGAG 3270

RESULT 12
US-09-429-516-2
; Sequence 2, Application US/09429516
; Patent No. 6251389
; GENERAL INFORMATION:
; APPLICANT: Magna, Holly
; APPLICANT: Schaffer, Paul
; APPLICANT: Lawton, Michael
; APPLICANT: Yocum, Sue
; APPLICANT: Mitchell, Peter
; APPLICANT: Hutchinson, Nancy
; APPLICANT: Murry, Lynn E.
; TITLE OF INVENTION: HUMAN NUCLEOTIDE PYROPHOSPHOS
; PHYDROLASE-2
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/429,516
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/996,083
; FILING DATE:
```

ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0420 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-855-0555  
TELEFAX: 650-845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4183 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: SATPF1008  
CLONE: 1388013  
US-09-429-516-2

Query Match 90.6%; Score 15.4; DB 3; Length 4183;  
Best Local Similarity 94.1%; Pred. No. 1.4e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGCCGAG 17  
|||||  
Db 3254 GCCACGCTTGCCGAG 3270

RESULT 13  
US-09-949-016-157184/c  
Sequence 157184, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 157184  
LENGTH: 601  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-157184

Query Match 84.7%; Score 14.4; DB 3; Length 601;  
Best Local Similarity 93.8%; Pred. No. 4.3e+02;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 CCCACGCTTGCCGAG 17  
|||||  
Db 71 CCCACTTGCCGAG 56

RESULT 14  
US-09-027-900-9/c  
Sequence 9, Application US/09027900  
Patent No. 6248551  
GENERAL INFORMATION:  
APPLICANT: De Reuse, Hilde  
APPLICANT: Skouloubri, Stephane  
APPLICANT: Labigne, Agnes  
TITLE OF INVENTION: HELICOBACTER ALIPHATIC AMIDASE AmiE  
TITLE OF INVENTION: POLYPEPTIDES, AND DNA SEQUENCES ENCODING THOSE  
TITLE OF INVENTION: POLYPEPTIDES

NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FINNEGAN, HENDERSON, FARABOW, GARRETT &  
ADDRESSEE: DUNNER, LLP  
STREET: 1300 I Street, NW  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/027,900  
FILING DATE: 23-FEB-1998  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/041,745  
FILING DATE: 28-MAR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Meyers, Kenneth J.  
REGISTRATION NUMBER: 25,146  
REFERENCE/DOCKET NUMBER: 03495-0165.00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1020 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1017  
US-09-027-900-9

Query Match 84.7%; Score 14.4; DB 3; Length 1020;  
Best Local Similarity 93.8%; Pred. No. 4.3e+02;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 CCCACGCTTGCCGAG 17  
|||||  
Db 980 CCCACGCTTGCCGAG 965

RESULT 15  
US-09-027-900-8/c  
Sequence 8, Application US/09027900  
Patent No. 6248551  
GENERAL INFORMATION:  
APPLICANT: De Reuse, Hilde  
APPLICANT: Skouloubri, Stephane  
APPLICANT: Labigne, Agnes  
TITLE OF INVENTION: HELICOBACTER ALIPHATIC AMIDASE AmiE  
TITLE OF INVENTION: POLYPEPTIDES, AND DNA SEQUENCES ENCODING THOSE  
TITLE OF INVENTION: POLYPEPTIDES  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FINNEGAN, HENDERSON, FARABOW, GARRETT &  
ADDRESSEE: DUNNER, LLP  
STREET: 1300 I Street, NW  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

```
;
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/027,900
; FILING DATE: 23-FEB-1998
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/041,745
; FILING DATE: 28-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 03495-0165.00000
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1520 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-027-900-8
;
; Query Match 84.7%; Score 14.4; DB 3; Length 1520;
; Best Local Similarity 93.8%; Pred. No. 4.3e+02;
; Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
;
Qy 2 CCCACGCTTGGCCGAG 17
Db 1454 CCCACGCTTGGCCGAG 1439
|||||
;
RESULT 16
US-09-027-900-7/c
; Sequence 7, Application US/09027900
; Patent No. 6248551
; GENERAL INFORMATION:
; APPLICANT: De Reuse, Hilde
; APPLICANT: Skouloubri, Stephanie
; APPLICANT: Labigne, Agnes
; TITLE OF INVENTION: HELICOBACTER ALIPHATIC AMIDASE Amie
; TITLE OF INVENTION: POLYPEPTIDES, AND DNA SEQUENCES ENCODING THOSE
; TITLE OF INVENTION: POLYPEPTIDES
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FINNEGAN, HENDERSON, FARABOW, GARRETT &
; ADDRESSEE: DUNNER, LLP
; STREET: 1300 I Street, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/027,900
; FILING DATE: 23-FEB-1998
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/041,745
; FILING DATE: 28-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 03495-0165.00000
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 7:
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1675 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-027-900-7
;
; Query Match 84.7%; Score 14.4; DB 3; Length 1675;
; Best Local Similarity 93.8%; Pred. No. 4.3e+02;
; Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
;
Qy 2 CCCACGCTTGGCCGAG 17
Db 1534 CCCACGCTTGGCCGAG 1519
|||||
;
RESULT 17
US-09-949-016-16127/c
; Sequence 16127, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16127
; LENGTH: 42242
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-16127
;
; Query Match 84.7%; Score 14.4; DB 3; Length 42242;
; Best Local Similarity 93.8%; Pred. No. 4.4e+02;
; Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
;
Qy 2 CCCACGCTTGGCCGAG 17
Db 7247 CCCACTCTTGGCCGAG 7232
|||||
;
RESULT 18
US-09-949-016-13514
; Sequence 13514, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13514
; LENGTH: 47698
; TYPE: DNA
```

```

; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1) - (47698)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13514

Query Match      84.7%; Score 14.4; DB 3; Length 47698;
Best Local Similarity 93.8%; Pred. No. 4.4e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCCACGCTTGCCGAG 17
Db 9925 CCCACGCTTGCTGAG 9940

RESULT 19
US-09-671-317-485/c
; Sequence 485, Application US/09671317
; Patent No. 6528260
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Chumakov, Ilya
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Cohen, Annick
; TITLE OF INVENTION: BIALLELIC MARKERS RELATED TO GENES INVOLVED IN DRUG METABOLISM
; FILE REFERENCE: 62 US3 CIP
; CURRENT APPLICATION NUMBER: US/09/671,317
; CURRENT FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US 09/536,178
; PRIOR FILING DATE: 2000-03-23
; PRIOR APPLICATION NUMBER: PCT/IB00/00403
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: US 60/126,269
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/131,961
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 977
; SOFTWARE: Patent.pm
; SEQ ID NO 485
; LENGTH: 49312
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 5466..7466
; OTHER INFORMATION: 5' regulatory region
; NAME/KEY: exon
; LOCATION: 7467..7725
; OTHER INFORMATION: exon 1
; NAME/KEY: exon
; LOCATION: 20256..20355
; OTHER INFORMATION: exon 2
; NAME/KEY: exon
; LOCATION: 36905..36975
; OTHER INFORMATION: exon 3
; NAME/KEY: exon
; LOCATION: 45167..45248
; OTHER INFORMATION: exon 4
; NAME/KEY: exon
; LOCATION: 45728..45965
; OTHER INFORMATION: exon 5
; NAME/KEY: misc feature
; LOCATION: 45966..49312
; OTHER INFORMATION: 3' regulatory region
; NAME/KEY: allele
; LOCATION: 7564
; OTHER INFORMATION: 10-286-289 : polymorphic base G or C
; NAME/KEY: allele
; LOCATION: 7619
; OTHER INFORMATION: 10-286-345 : polymorphic base A or T
; NAME/KEY: allele
; LOCATION: 7649
; OTHER INFORMATION: 10-286-345.mis complement

; OTHER INFORMATION: 10-286-375 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 17258
; OTHER INFORMATION: 12-425-57 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 21590
; OTHER INFORMATION: 12-421-135 : insertion of T
; NAME/KEY: allele
; LOCATION: 21595
; OTHER INFORMATION: 12-421-140 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 36971
; OTHER INFORMATION: 10-523-232 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 45214
; OTHER INFORMATION: 10-289-201 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 45741
; OTHER INFORMATION: 10-290-37 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 46029
; OTHER INFORMATION: 10-290-326 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 46032
; OTHER INFORMATION: 10-290-328 : deletion of G
; NAME/KEY: primer bind
; LOCATION: 7276..7294
; OTHER INFORMATION: 10-286.pu
; NAME/KEY: primer bind
; LOCATION: 7676..7694
; OTHER INFORMATION: 10-286.rp complement
; NAME/KEY: primer bind
; LOCATION: 16839..16856
; OTHER INFORMATION: 12-425.rp
; NAME/KEY: primer bind
; LOCATION: 17297..17314
; OTHER INFORMATION: 12-425.pu complement
; NAME/KEY: primer bind
; LOCATION: 21456..21474
; OTHER INFORMATION: 12-421.pu
; NAME/KEY: primer bind
; LOCATION: 21886..21906
; OTHER INFORMATION: 12-421.rp complement
; NAME/KEY: primer bind
; LOCATION: 36740..36758
; OTHER INFORMATION: 10-523.pu
; NAME/KEY: primer bind
; LOCATION: 36997..37015
; OTHER INFORMATION: 10-523.rp complement
; NAME/KEY: primer bind
; LOCATION: 45020..45037
; OTHER INFORMATION: 10-289.pu
; NAME/KEY: primer bind
; LOCATION: 45413..45432
; OTHER INFORMATION: 10-289.rp complement
; NAME/KEY: primer bind
; LOCATION: 45705..45724
; OTHER INFORMATION: 10-290.pu
; NAME/KEY: primer bind
; LOCATION: 46104..46123
; OTHER INFORMATION: 10-290.rp complement
; NAME/KEY: primer bind
; LOCATION: 7545..7563
; OTHER INFORMATION: 10-286-289.mis
; NAME/KEY: primer bind
; LOCATION: 7565..7583
; OTHER INFORMATION: 10-286-289.mis complement
; NAME/KEY: primer bind
; LOCATION: 7600..7618
; OTHER INFORMATION: 10-286-345.mis
; NAME/KEY: primer bind
; LOCATION: 7620..7638
; OTHER INFORMATION: 10-286-345.mis complement
```

NAME/KEY: primer\_bind  
LOCATION: 7630..7648  
OTHER INFORMATION: 10-286-375.mis  
NAME/KEY: primer\_bind  
LOCATION: 7650..7668  
OTHER INFORMATION: 10-286-375.mis complement  
NAME/KEY: primer\_bind  
LOCATION: 17239..17257  
OTHER INFORMATION: 12-425-57.mis  
NAME/KEY: primer\_bind  
LOCATION: 17259..17277  
OTHER INFORMATION: 12-425-57.mis complement  
NAME/KEY: primer\_bind  
LOCATION: 21576..21594  
OTHER INFORMATION: 12-421-140.mis  
NAME/KEY: primer\_bind  
LOCATION: 21596..21614  
OTHER INFORMATION: 12-421-140.mis complement  
NAME/KEY: primer\_bind  
LOCATION: 36952..36970  
OTHER INFORMATION: 10-523-232.mis  
NAME/KEY: primer\_bind  
LOCATION: 36972..36990  
OTHER INFORMATION: 10-523-232.mis complement  
NAME/KEY: primer\_bind  
LOCATION: 45195..45213  
OTHER INFORMATION: 10-289-201.mis  
NAME/KEY: primer\_bind  
LOCATION: 45215..45233  
OTHER INFORMATION: 10-289-201.mis complement  
NAME/KEY: primer\_bind  
LOCATION: 45722..45740  
OTHER INFORMATION: 10-290-37.mis  
NAME/KEY: primer\_bind  
LOCATION: 45742..45760  
OTHER INFORMATION: 10-290-37.mis complement  
NAME/KEY: primer\_bind  
LOCATION: 46010..46028  
OTHER INFORMATION: 10-290-326.mis  
NAME/KEY: primer\_bind  
LOCATION: 46030..46048  
OTHER INFORMATION: 10-290-326.mis complement  
NAME/KEY: misc\_binding  
LOCATION: 7552..7576  
OTHER INFORMATION: 10-286-289.probe  
NAME/KEY: misc\_binding  
LOCATION: 7607..7631  
OTHER INFORMATION: 10-286-345.probe  
NAME/KEY: misc\_binding  
LOCATION: 7637..7661  
OTHER INFORMATION: 10-286-375.probe  
NAME/KEY: misc\_binding  
LOCATION: 17245..17270  
OTHER INFORMATION: 12-425-57.probe  
NAME/KEY: misc\_binding  
LOCATION: 21583..21607  
OTHER INFORMATION: 12-421-140.probe  
NAME/KEY: misc\_binding  
LOCATION: 36959..36983  
OTHER INFORMATION: 10-523-232.probe  
NAME/KEY: misc\_binding  
LOCATION: 45202..45226  
OTHER INFORMATION: 10-289-201.probe  
NAME/KEY: misc\_binding  
LOCATION: 45729..45753  
OTHER INFORMATION: 10-290-37.probe  
NAME/KEY: misc\_binding  
LOCATION: 46017..46041  
OTHER INFORMATION: 10-290-326.probe  
US-09-671-317-485

Query Match 84.7%; Score 14.4; DB 3; Length 49312;  
Best Local Similarity 93.8%; Pred. No. 4.4e+02;

Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 2 CCCACGCTTGCCGAG 17  
|||||  
Db 12970 CCCACTCTTGCCGAG 12955

RESULT 20

US-09-949-016-16803/c  
Sequence 16803, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 16803  
LENGTH: 67002  
TYPE: DNA  
ORGANISM: Human  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)..(67002)  
OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-16803

Query Match 84.7%; Score 14.4; DB 3; Length 67002;  
Best Local Similarity 93.8%; Pred. No. 4.4e+02;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 CCCACGCTTGCCGAG 17  
|||||  
Db 41452 CCCAGGCTTGCCGAG 41437

RESULT 21

US-09-332-522E-17/c  
Sequence 17, Application US/09332522E  
Patent No. 6781028  
GENERAL INFORMATION:  
APPLICANT: Costa, M.  
APPLICANT: Doberstein, S.  
APPLICANT: Elson, S.  
APPLICANT: Ferguson, K.  
APPLICANT: Homberger, S.  
TITLE OF INVENTION: ANIMAL MODELS AND METHODS FOR ANALYSIS OF LIPID METABOLISM AND S  
FILE REFERENCE: 7326-101, EX99-004  
CURRENT APPLICATION NUMBER: US/09/332,522E  
CURRENT FILING DATE: 1999-06-14  
NUMBER OF SEQ ID NOS: 95  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 17  
LENGTH: 20  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Oligonucleotide  
US-09-332-522E-17

Query Match 81.2%; Score 13.8; DB 3; Length 20;  
Best Local Similarity 88.2%; Pred. No. 8.6e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCCCAGCTTGGCCGAG 17  
|||||  
Db 18 GCCCAGCATGACCGAG 2

RESULT 22  
; Sequence 2426, Application US/09422978  
; Patent No. 6537751  
; GENERAL INFORMATION:  
; APPLICANT: Cohen, Daniel  
; APPLICANT: Blumenfeld, Marta  
; APPLICANT: Chumakov, Ilya  
; TITLE OF INVENTION: Biallelic markers for use in constructing a high density...  
; FILE REFERENCE: GENSET.020CPI  
; CURRENT APPLICATION NUMBER: US/09/422,978  
; CURRENT FILING DATE: 1999-10-20  
; EARLIER APPLICATION NUMBER: US 09/298,850  
; EARLIER FILING DATE: 1999-04-21  
; EARLIER APPLICATION NUMBER: US 60/109,732  
; EARLIER FILING DATE: 1998-11-23  
; EARLIER APPLICATION NUMBER: US 60/082,614  
; EARLIER FILING DATE: 1998-04-21  
; NUMBER OF SEQ ID NOS: 11796  
; SEQ ID NO 2426  
; LENGTH: 47  
; TYPE: DNA  
; ORGANISM: Homo Sapiens  
; FEATURE:  
; NAME/KEY: allele  
; LOCATION: 24  
; OTHER INFORMATION: 99-10951-434 : polymorphic base T or C  
US-09-422-978-2426

Query Match 81.2%; Score 13.8; DB 3; Length 47;  
Best Local Similarity 88.2%; Pred. No. 8.6e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCCCAGCTTGGCCGAG 17  
|||||  
Db 27 GCCCAGCTTGTACAG 43

RESULT 23  
; Sequence 2571, Application US/09313294A  
; Patent No. 6476212  
; GENERAL INFORMATION:  
; APPLICANT: Lalgudi, Raghunath V.  
; APPLICANT: Ito, Laura Y.  
; APPLICANT: Sherman, Bradley K.  
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR  
; FILE REFERENCE: PL-0017 US  
; CURRENT APPLICATION NUMBER: US/09/313,294A  
; CURRENT FILING DATE: 1999-05-14  
; NUMBER OF SEQ ID NOS: 7600  
; SOFTWARE: PERL Program  
; SEQ ID NO 2571  
; LENGTH: 283  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: Incyte ID No. 6476212 700552904H1  
US-09-313-294A-2571

Query Match 81.2%; Score 13.8; DB 3; Length 283;  
Best Local Similarity 88.2%; Pred. No. 8.7e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCCCAGCTTGGCCGAG 17  
|||||

Db 106 GCCCACCCTTGCCCAAG 90  
  
RESULT 24  
; Sequence 5109, Application US/09313294A  
; Patent No. 6476212  
; GENERAL INFORMATION:  
; APPLICANT: Lalgudi, Raghunath V.  
; APPLICANT: Ito, Laura Y.  
; APPLICANT: Sherman, Bradley K.  
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR  
; FILE REFERENCE: PL-0017 US  
; CURRENT APPLICATION NUMBER: US/09/313,294A  
; CURRENT FILING DATE: 1999-05-14  
; NUMBER OF SEQ ID NOS: 7600  
; SOFTWARE: PERL Program  
; SEQ ID NO 5109  
; LENGTH: 293  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: Incyte ID No. 6476212 700349603H1  
; NAME/KEY: unsure  
; LOCATION: 232, 236, 244, 250, 278, 280, 284  
; OTHER INFORMATION: a, t, c, g, or other  
US-09-313-294A-5109

Query Match 81.2%; Score 13.8; DB 3; Length 293;  
Best Local Similarity 88.2%; Pred. No. 8.7e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCCCAGCTTGGCCGAG 17  
|||||  
Db 180 GCCCACCCTTGCCCAAG 164

RESULT 25  
; Sequence 44, Application US/08990823D  
; Patent No. 6228371  
; GENERAL INFORMATION:  
; APPLICANT: Nano, Francis  
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding  
; FILE REFERENCE: 49086  
; CURRENT APPLICATION NUMBER: US/08/990,823D  
; CURRENT FILING DATE: 1997-12-15  
; EARLIER APPLICATION NUMBER: US 96/10375  
; EARLIER FILING DATE: 1996-06-14  
; EARLIER APPLICATION NUMBER: 60/000,254  
; EARLIER FILING DATE: 1995-06-15  
; NUMBER OF SEQ ID NOS: 113  
; SOFTWARE: PatentIN Ver. 2.0  
; SEQ ID NO 44  
; LENGTH: 385  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
US-08-990-823-44

Query Match 81.2%; Score 13.8; DB 3; Length 385;  
Best Local Similarity 88.2%; Pred. No. 8.7e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCCCAGCTTGGCCGAG 17  
|||||  
Db 11 GCCCAGCTTGGCCTAG 27

RESULT 26  
; Sequence 44, Application US/09477135A

```
; Patent No. 6572865
; GENERAL INFORMATION:
; APPLICANT: Nano, Francis
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
; TITLE OF INVENTION: immunostimulatory Peptides
; FILE REFERENCE: 52888
; CURRENT APPLICATION NUMBER: US/09/477,135A
; PRIOR FILING DATE: 2000-01-03
; PRIOR APPLICATION NUMBER: 08990823
; PRIOR FILING DATE: 1997-12-15
; PRIOR APPLICATION NUMBER: US 96/10375
; PRIOR FILING DATE: 1996-06-14
; PRIOR APPLICATION NUMBER: 60/000,254
; PRIOR FILING DATE: 1995-06-15
; NUMBER OF SEQ ID NOS: 169
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 44
; LENGTH: 385
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-09-477-135A-44

Query Match      81.2%; Score 13.8; DB 3; Length 385;
Best Local Similarity 88.2%; Pred. No. 8.7e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 GCCCAGCGTTGCCGCGAG 17
      ||||| ||||| ||||| |||||
Db      11 GCCCGCGCTTGCCCTAG 27

RESULT 27
US-10-233-942-3
; Sequence 3, Application US/10233942
; Patent No. 697178
; GENERAL INFORMATION:
; APPLICANT: Hodge, Timothy et al
; TITLE OF INVENTION: System for Automated Transgenic Screening
; FILE REFERENCE: 023131.41500
; CURRENT APPLICATION NUMBER: US/10/233,942
; CURRENT FILING DATE: 2002-09-03
; PRIOR APPLICATION NUMBER: 60/230,371
; PRIOR FILING DATE: 2000-06-09
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 600
; TYPE: DNA
; ORGANISM: streptomyces alboniger
US-10-233-942-3

Query Match      81.2%; Score 13.8; DB 4; Length 600;
Best Local Similarity 88.2%; Pred. No. 8.7e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 GCCCAGCGTTGCCGCGAG 17
      ||||| ||||| ||||| |||||
Db      275 GCCCGCGCATGCCGCGAG 291

RESULT 28
US-09-945-952A-3
; Sequence 3, Application US/09945952A
; Patent No. 7011943
; GENERAL INFORMATION:
; APPLICANT: Hodge, Timothy A.
; TITLE OF INVENTION: System for Automated Transgenic Screening
; FILE REFERENCE: 023131.41500
; CURRENT APPLICATION NUMBER: US/09/945,952A
; CURRENT FILING DATE: 2001-12-06
; PRIOR APPLICATION NUMBER: 60/230,371
; PRIOR FILING DATE: 2000-06-09
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 600
; TYPE: DNA
; ORGANISM: streptomyces alboniger
US-09-945-952A-3

; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 600
; TYPE: DNA
; ORGANISM: streptomyces alboniger
US-09-945-952A-3

Query Match      81.2%; Score 13.8; DB 5; Length 600;
Best Local Similarity 88.2%; Pred. No. 8.7e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 GCCCAGCGTTGCCGCGAG 17
      ||||| ||||| ||||| |||||
Db      275 GCCCGCGCATGCCGCGAG 291

RESULT 29
US-09-949-016-121284
; Sequence 121284, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 121284
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-121284

Query Match      81.2%; Score 13.8; DB 3; Length 601;
Best Local Similarity 88.2%; Pred. No. 8.7e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 GCCCAGCGTTGCCGCGAG 17
      ||||| ||||| ||||| |||||
Db      351 GCCCATGCTTGCCCAAG 367

RESULT 30
US-09-949-016-205926
; Sequence 205926, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 205926
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-205926
```



US-09-949-016-205926

Query Match 81.2%; Score 13.8; DB 3; Length 601;  
Best Local Similarity 88.2%; Pred. No. 8.7e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCCCAGCTTGGCCGAG 17  
||||| ||||| |||

Db 25 GCCCAGCTTGGCCGAG 41

RESULT 31  
US-09-949-016-205927  
; Sequence 205927, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 205927  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-205927

Query Match 81.2%; Score 13.8; DB 3; Length 601;  
Best Local Similarity 88.2%; Pred. No. 8.7e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCCCAGCTTGGCCGAG 17  
||||| ||||| |||

Db 52 GCCCAGCTTGGCCGAG 68

RESULT 32  
US-09-902-540-5464/c  
; Sequence 5464, Application US/09902540  
; Patent No. 6833447  
; GENERAL INFORMATION:  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Wiegand, Roger C.  
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(15849)B  
; CURRENT APPLICATION NUMBER: US/09/902,540  
; CURRENT FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: 60/217,883  
; PRIOR FILING DATE: 2000-07-10  
; NUMBER OF SEQ ID NOS: 16825  
; SEQ ID NO 5464  
; LENGTH: 795  
; TYPE: DNA  
; ORGANISM: Myxococcus xanthus  
US-09-902-540-5464

Query Match 81.2%; Score 13.8; DB 3; Length 795;  
Best Local Similarity 88.2%; Pred. No. 8.7e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCCCAGCTTGGCCGAG 17  
||||| ||||| |||

Db 712 GCCCAGCTTGGCCGAG 696

RESULT 33  
US-09-252-991A-6261/c  
; Sequence 6261, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 6261  
; LENGTH: 999  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-6261

Query Match 81.2%; Score 13.8; DB 3; Length 999;  
Best Local Similarity 88.2%; Pred. No. 8.7e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCCCAGCTTGGCCGAG 17  
||||| ||||| |||||

Db 343 GCCCAGCTTGGCCGAG 327

RESULT 34  
US-09-252-991A-547/c  
; Sequence 547, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 547  
; LENGTH: 1176  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-547

Query Match 81.2%; Score 13.8; DB 3; Length 1176;  
Best Local Similarity 88.2%; Pred. No. 8.7e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCCCAGCTTGGCCGAG 17  
||||| ||||| |||||

Db 668 GCCCAGCTTGGCCGAG 652

RESULT 35  
US-09-540-236-266  
; Sequence 266, Application US/09540236  
; Patent No. 6673910  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR  
; FILE REFERENCE: 107196.136

; FILE REFERENCE: 2709.2005-001  
; CURRENT APPLICATION NUMBER: US/09/540.236  
; CURRENT FILING DATE: 2000-04-04  
; NUMBER OF SEQ ID NOS: 3840  
; SEQ ID NO 266  
; LENGTH: 1182  
; TYPE: DNA  
; ORGANISM: M.catarrhalis  
US-09-540-236-266

Query Match 81.2%; Score 13.8; DB 3; Length 1182;  
Best Local Similarity 88.2%; Pred. No. 8.7e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCCGAG 17  
||||| ||||| |||||  
Db 1080 GCCCAAGCTTAGCCGAG 1096

RESULT 36  
US-08-095-726-13  
; Sequence 13, Application US/08095726  
; Patent No. 5530188  
; GENERAL INFORMATION:  
; APPLICANT: Ausich, Rodney L  
; APPLICANT: Brinkhaus, Friedhelm L  
; APPLICANT: Mukharji, Indrani  
; APPLICANT: Proffitt, John H  
; APPLICANT: Yarger, James G  
; APPLICANT: Yen, Hwei-Che B  
; TITLE OF INVENTION: Beta-Carotene Biosynthesis in  
; Genetically Engineered Hosts  
; NUMBER OF SEQUENCES: 79  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Amoco Corp., Patents and Licensing Dept  
; STREET: 200 E Randolph St  
; CITY: Chicago  
; STATE: IL  
; COUNTRY: USA  
; ZIP: 60680-0703  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.24  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/095,726  
; FILING DATE: 21-JUL-1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; FILING DATE: 30-OCT-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Galloway, No. 5530188val B  
; TELEPHONE: 3128567180  
; TELEFAX: 3128564972  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1235 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-095-726-13

Query Match 81.2%; Score 13.8; DB 2; Length 1235;  
Best Local Similarity 88.2%; Pred. No. 8.7e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 GCCCAGCTTGGCCGAG 17  
||||| ||||| |||||  
Db 209 GCCCAGCTTGGCCGG 225

Query Match 81.2%; Score 13.8; DB 2; Length 1235;  
Best Local Similarity 88.2%; Pred. No. 8.7e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCCGAG 17  
||||| ||||| |||||  
Db 209 GCCCAGCTTGGCCGG 225

RESULT 37  
US-08-095-726-15  
; Sequence 15, Application US/08095726  
; Patent No. 5530188  
; GENERAL INFORMATION:  
; APPLICANT: Ausich, Rodney L  
; APPLICANT: Brinkhaus, Friedhelm L  
; APPLICANT: Mukharji, Indrani  
; APPLICANT: Proffitt, John H  
; APPLICANT: Yarger, James G  
; APPLICANT: Yen, Hwei-Che B  
; TITLE OF INVENTION: Beta-Carotene Biosynthesis in  
; Genetically Engineered Hosts  
; NUMBER OF SEQUENCES: 79  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Amoco Corp., Patents and Licensing Dept  
; STREET: 200 E Randolph St  
; CITY: Chicago  
; STATE: IL  
; COUNTRY: USA  
; ZIP: 60680-0703  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.24  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/095,726  
; FILING DATE: 21-JUL-1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; FILING DATE: 30-OCT-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Galloway, No. 5530188val B  
; TELEPHONE: 3128567180  
; TELEFAX: 3128564972  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1235 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-095-726-15

Query Match 81.2%; Score 13.8; DB 2; Length 1235;  
Best Local Similarity 88.2%; Pred. No. 8.7e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCCGAG 17  
||||| ||||| |||||  
Db 209 GCCCAGCTTGGCCGG 225

RESULT 38  
US-08-096-623A-13  
; Sequence 13, Application US/08096623A  
; Patent No. 5684238  
; GENERAL INFORMATION:  
; APPLICANT: Ausich, Rodney L  
; APPLICANT: Brinkhaus, Friedhelm L  
; APPLICANT: Mukharji, Indrani  
; APPLICANT: Proffitt, John H  
; APPLICANT: Yarger, James G  
; APPLICANT: Yen, Hwei-Che B  
; TITLE OF INVENTION: Biosynthesis of Zeaxanthin and  
; Glycosylated Zeaxanthin in Genetically Engineered Hosts  
; NUMBER OF SEQUENCES: 104  
; CORRESPONDENCE ADDRESS:

ADDRESSEE: Welsh & Katz, Ltd.  
STREET: 120 S. Riverside Plaza, 22nd Floor  
CITY: Chicago  
STATE: IL  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/096,623A  
FILING DATE: 22-JUL-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/805,061  
FILING DATE: 09-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/662,921  
FILING DATE: 28-FEB-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/562,674  
FILING DATE: 03-AUG-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/525,551  
FILING DATE: 18-MAY-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/487,613  
FILING DATE: 02-MAR-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Gamson, Edward P.  
REGISTRATION NUMBER: 29,381  
REFERENCE/DOCKET NUMBER: AMO-006.1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 655-1500  
TELEFAX: (312) 655-1501  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1235 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
POSITION IN GENOME:  
MAP POSITION: -19 to 1216  
UNITS: bp  
US-08-096-623A-13

Query Match 81.2%; Score 13.8; DB 2; Length 1235;  
Best Local Similarity 88.2%; Pred. No. 8.7e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCCCAGCTTGGCCGAG 17  
|||||  
DB 209 GCCCAGCTTGGCCGAG 225

## RESULT 39

US-08-096-623A-15  
Sequence 15, Application US/08096623A  
Patent No. 5684238  
GENERAL INFORMATION:  
APPLICANT: Ausich, Rodney L.  
APPLICANT: Brinkhaus, Friedhelm L.  
APPLICANT: Mukharji, Indrani  
APPLICANT: Proffitt, John H.  
APPLICANT: Yarger, James G.  
APPLICANT: Yen, Hwei-Che B.  
TITLE OF INVENTION: Biosynthesis of Zeaxanthin and  
NUMBER OF SEQUENCES: 104  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Welsh & Katz, Ltd.  
STREET: 120 S. Riverside Plaza, 22nd Floor  
CITY: Chicago  
STATE: IL  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/096,623A  
FILING DATE: 22-JUL-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/805,061  
FILING DATE: 09-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/662,921  
FILING DATE: 28-FEB-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/562,674  
FILING DATE: 03-AUG-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/525,551  
FILING DATE: 18-MAY-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/487,613  
FILING DATE: 02-MAR-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Gamson, Edward P.  
REGISTRATION NUMBER: 29,381  
REFERENCE/DOCKET NUMBER: AMO-006.1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 655-1500  
TELEFAX: (312) 655-1501  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1235 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
POSITION IN GENOME:  
MAP POSITION: -19 to 1216  
UNITS: bp  
US-08-096-623A-15

Query Match 81.2%; Score 13.8; DB 2; Length 1235;  
Best Local Similarity 88.2%; Pred. No. 8.7e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCCCAGCTTGGCCGAG 17  
|||||  
DB 209 GCCCAGCTTGGCCGAG 225

## RESULT 40

US-09-252-991A-1019/c  
Sequence 1019, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27

```
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1019
; LENGTH: 1251
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1019

Query Match      81.2%; Score 13.8; DB 3; Length 1251;
Best Local Similarity 88.2%; Pred. No. 8.7e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 GCCCAGCGTTGCCCGAG 17
        ||||| ||||| |||||
Db      225 GCCCAGCGCTGGCGGTG 209

RESULT 41
US-08-681-129-1/c
; Sequence 1, Application US/08681129
; Patent No. 5738854
; GENERAL INFORMATION:
; APPLICANT: Mettenleiter, Thomas Cristoph
; TITLE OF INVENTION: Pseudorabies virus vaccine
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Organon Teknika Corporation
; STREET: 1330 Piccard Drive
; CITY: Rockville
; STATE: Maryland
; COUNTRY: U.S.A.
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; APPLICATION DATA:
; APPLICATION NUMBER: US/08/681,129
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/244,446
; FILING DATE: 02-JUN-1994
; CLASSIFICATION: 424
; APPLICATION NUMBER: EP 92.203.079.6
; FILING DATE: 06-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Gornley, Mary E.
; REGISTRATION NUMBER: 34,409
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 258-5200
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1578 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Pseudorabies virus
; STRAIN: Kaplan
; US-08-681-129-1

Query Match      81.2%; Score 13.8; DB 2; Length 1578;
Best Local Similarity 88.2%; Pred. No. 8.7e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 GCCCAGCGTTGCCCGAG 17
        ||||| ||||| |||||
Db      227 GCCCAGCGCGGCCGAG 211

RESULT 42
US-09-252-991A-569
; Sequence 569, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 569
; LENGTH: 1608
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-569

Query Match      81.2%; Score 13.8; DB 3; Length 1608;
Best Local Similarity 88.2%; Pred. No. 8.7e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 GCCCAGCGTTGCCCGAG 17
        ||||| ||||| |||||
Db      1334 GCCCAGCGTTGCCCGAG 1350

RESULT 43
US-09-869-588-32
; Sequence 32, Application US/09869588
; Patent No. 6790657
; GENERAL INFORMATION:
; APPLICANT: Arya
; TITLE OF INVENTION: Lentivirus Vector System
; FILE REFERENCE: 59316
; CURRENT APPLICATION NUMBER: US/09/869,588
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: PCT/US00/00390
; PRIOR FILING DATE: 2000-01-06
; PRIOR APPLICATION NUMBER: 60/115,247
; PRIOR FILING DATE: 1999-01-07
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 32
; LENGTH: 1616
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: IRES and
; OTHER INFORMATION: Puromycin sequences
US-09-869-588-32

Query Match      81.2%; Score 13.8; DB 3; Length 1616;
Best Local Similarity 88.2%; Pred. No. 8.7e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 GCCCAGCGTTGCCCGAG 17
        ||||| ||||| |||||
Db      1244 GCCCGCGCATGCCCGAG 1260

RESULT 44
US-09-252-991A-1111
; Sequence 1111, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
```

; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 1111  
; LENGTH: 1653  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-1111

Query Match 81.2%; Score 13.8; DB 3; Length 1653;  
Best Local Similarity 88.2%; Pred. No. 8.7e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCCCAGCTTGGCCGAG 17  
|||||  
Db 252 GCCCAGCTTGGCCGTG 268

## RESULT 45

US-09-165-042-2  
; Sequence 2, Application US/09165042  
; Patent No. 6100077  
; GENERAL INFORMATION:  
; APPLICANT: Sturley, Stephen L.  
; APPLICANT: Oelkers, Peter  
; TITLE OF INVENTION: ISOLATION OF A GENE ENCODING DIACYLGLYCEROL  
; FILE REFERENCE: 0575/56331  
; CURRENT APPLICATION NUMBER: US/09/165,042  
; CURRENT FILING DATE: 1998-10-01  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 1976  
; TYPE: DNA  
; ORGANISM: Yeast  
US-09-165-042-2

Query Match 81.2%; Score 13.8; DB 3; Length 1976;  
Best Local Similarity 88.2%; Pred. No. 8.8e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCCCAGCTTGGCCGAG 17  
|||||  
Db 208 GCCCAGCTTGGCTCG 224

## RESULT 46

US-09-252-991A-531/c  
; Sequence 531, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 531  
; LENGTH: 1983  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-531

Query Match 81.2%; Score 13.8; DB 3; Length 1983;  
Best Local Similarity 88.2%; Pred. No. 8.8e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCCCAGCTTGGCCGAG 17  
|||||  
Db 693 GCCCAGCTTGGCCGAG 677

## RESULT 47

US-09-252-991A-6426/c  
; Sequence 6426, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 6426  
; LENGTH: 2049  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1174), (1177)  
; OTHER INFORMATION: Identity of nucleotide at the above locations are unknown.  
US-09-252-991A-6426

Query Match 81.2%; Score 13.8; DB 3; Length 2049;  
Best Local Similarity 88.2%; Pred. No. 8.8e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCCCAGCTTGGCCGAG 17  
|||||  
Db 1520 GCCCAGCTTGGCCGAG 1504

## RESULT 48

US-08-358-117-1/c  
; Sequence 1, Application US/08358117  
; Patent No. 5608147  
; GENERAL INFORMATION:  
; APPLICANT: Kaphammer, Bryan J.  
; TITLE OF INVENTION: ttda Gene Selectable Markers in Plants and the  
; TITLE OF INVENTION: Use Thereof  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; STREET: 1100 New York Avenue, Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/358,117  
; FILING DATE: 15-DEC-1994  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Esmond, Robert W.  
; REGISTRATION NUMBER: 32,893  
; REFERENCE/DOCKET NUMBER: 1405.0030001

; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 371-2600  
; TELEFAX: (202) 371-2540  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2058 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 751..1611  
; US-08-358-117-1

Query Match 81.2%; Score 13.8; DB 2; Length 2058;  
Best Local Similarity 88.2%; Pred. No. 8.8e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17  
Db 1573 GCTCAGCGTTGGCCGAG 1557

## RESULT 49

US-08-470-588-1/c  
; Sequence 1, Application US/08470588  
; Patent No. 6100446  
; GENERAL INFORMATION:  
; APPLICANT: STREIBER, WOLFGANG R.  
; APPLICANT: TIMMIS, KENNETH N.  
; APPLICANT: ZENK, MEINHART H.  
; TITLE OF INVENTION: MICROORGANISMS AND PLASMIDS FOR  
; TITLE OF INVENTION: 2,4-DICHLOROPHENOXACETIC ACID (2,4-D) MONOOXYGENASE  
; TITLE OF INVENTION: FORMATION AND PROCESS FOR THE PRODUCTION OF THESE PLASMIDS  
; TITLE OF INVENTION: AND STRAINS  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FROMMER LAWRENCE & HAUG LLP  
; STREET: 745 FIFTH AVENUE  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: USA  
; ZIP: 10151

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/470,588  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: LAWRENCE, WILLIAM F.  
; REGISTRATION NUMBER: 28,029  
; REFERENCE/DOCKET NUMBER: 514413-3526  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-588-0800  
; TELEFAX: 212-588-0500

INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2058 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; US-08-470-588-1

Query Match 81.2%; Score 13.8; DB 3; Length 2058;  
Best Local Similarity 88.2%; Pred. No. 8.8e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17

Db 1573 GCTCAGCGTTGGCCGAG 1557

## RESULT 50

US-09-976-594-315/c  
; Sequence 315, Application US/09976594  
; Patent No. 6673549  
; GENERAL INFORMATION:  
; APPLICANT: Furness, Michael  
; APPLICANT: Buchbinder, Jenny  
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS  
; FILE REFERENCE: PA-0041 US  
; CURRENT APPLICATION NUMBER: US/09/976,594  
; PRIOR FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: 60/240,409  
; PRIOR FILING DATE: 2000-10-12  
; NUMBER OF SEQ ID NOS: 1143  
; SOFTWARE: PERL Program  
; SEQ ID NO 315  
; LENGTH: 2169  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: incyte ID No. 6673549 223407.4  
; NAME/KEY: unsure  
; LOCATION: 1034, 1043  
; OTHER INFORMATION: a, t, c, g, or other  
; US-09-976-594-315

Query Match 81.2%; Score 13.8; DB 3; Length 2169;  
Best Local Similarity 88.2%; Pred. No. 8.8e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17  
Db 268 GCCCAGCGTTGGCCGGG 252

## RESULT 51

US-09-613-444-1  
; Sequence 1, Application US/09613444  
; Patent No. 644427  
; GENERAL INFORMATION:  
; APPLICANT: Ludwig, Erwin H.  
; APPLICANT: Farese, Robert V.  
; APPLICANT: Innerarity, Thomas L.  
; APPLICANT: Cases, Sylvaine  
; TITLE OF INVENTION: Polymorphisms in a Diacylglycerol  
; TITLE OF INVENTION: Acyltransferase Gene, and Methods of Use Thereof.  
; FILE REFERENCE: 6510191US1  
; CURRENT APPLICATION NUMBER: US/09/613,444  
; CURRENT FILING DATE: 2000-07-11  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 2174  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; US-09-613-444-1

Query Match 81.2%; Score 13.8; DB 3; Length 2174;  
Best Local Similarity 88.2%; Pred. No. 8.8e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17  
Db 1684 GCCCAGCGTTGGCTGCG 1700

## RESULT 52

US-10-142-516-1

; Sequence 1, Application US/10142516  
; Patent No. 6911539  
; GENERAL INFORMATION:  
; APPLICANT: Ludwig, Erwin H.  
; APPLICANT: Fareese, Robert V.  
; APPLICANT: Innerarity, Thomas L.  
; APPLICANT: Cases, Sylvaine  
; TITLE OF INVENTION: Polymorphisms in a Diacylglycerol  
; FILE REFERENCE: UCAL-191CON  
; CURRENT APPLICATION NUMBER: US/10/142,516  
; CURRENT FILING DATE: 2002-05-08  
; PRIOR APPLICATION NUMBER: 09/613,444  
; PRIOR FILING DATE: 2000-07-11  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 2174  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-142-516-1

Query Match 81.2%; Score 13.8; DB 3; Length 2174;  
Best Local Similarity 88.2%; Pred. No. 8.8e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 GCCCAGCTTGCCGCG 17  
DB 1684 GCCCAGCTTGCGTGG 1700

RESULT 53  
US-09-902-540-5116  
; Sequence 5116, Application US/09902540  
; Patent No. 6833447  
; GENERAL INFORMATION:  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Wiegand, Roger C.  
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(15849)B  
; CURRENT APPLICATION NUMBER: US/09/902,540  
; CURRENT FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: 60/217,883  
; PRIOR FILING DATE: 2000-07-10  
; NUMBER OF SEQ ID NOS: 16825  
; SEQ ID NO 5116  
; LENGTH: 2184  
; TYPE: DNA  
; ORGANISM: Myxococcus xanthus  
US-09-902-540-5116

Query Match 81.2%; Score 13.8; DB 3; Length 2184;  
Best Local Similarity 88.2%; Pred. No. 8.8e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 GCCCAGCTTGCCGCG 17  
DB 1610 GCTCAGCGTGCCGCG 1626

RESULT 54  
US-09-252-991A-1036  
; Sequence 1036, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 1036  
; LENGTH: 2847  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-1036

Query Match 81.2%; Score 13.8; DB 3; Length 2847;  
Best Local Similarity 88.2%; Pred. No. 8.8e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 GCCCAGCTTGCCGCG 17  
DB 2660 GCCCAGCGTGCCGCGTG 2676

RESULT 55  
US-08-162-809-1  
; Sequence 1, Application US/08162809  
; Patent No. 5457048  
; GENERAL INFORMATION:  
; APPLICANT: Pasquale, Elena B.  
; APPLICANT: Sajjadi, Fereydoun G.  
; TITLE OF INVENTION: NOVEL EPH-RELATED TYROSINE KINASES,  
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES, AND METHODS OF USE  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CAMPBELL AND FLORES  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: United States of America  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/162,809  
; FILING DATE:  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn A.  
; REGISTRATION NUMBER: 31,815  
; REFERENCE/DOCKET NUMBER: P-LJ 9503  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3133 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: both  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: join(3..419, 421..2858)  
US-08-162-809-1

Query Match 81.2%; Score 13.8; DB 2; Length 3133;  
Best Local Similarity 88.2%; Pred. No. 8.8e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 GCCCAGCTTGCCGCG 17  
DB 2507 GGCCTCGCTTGCCGCG 2523

```

RESULT 56
US-09-902-540-4713/c
; Sequence 4713, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,893
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 4713
; LENGTH: 3453
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
; US-09-902-540-4713

Query Match      81.2%; Score 13.8; DB 3; Length 3453;
Best Local Similarity 88.2%; Pred. No. 8.8e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GCCCAGCGCTGGCGGAG 17
Db 2623 GCCCAGCGCTGGCGGAG 2607
|||||

RESULT 57
US-10-094-749-957
; Sequence 957, Application US/10094749
; Patent No. 6979557
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOTYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 957
; LENGTH: 3636
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-094-749-957

Query Match      81.2%; Score 13.8; DB 4; Length 3636;
Best Local Similarity 88.2%; Pred. No. 8.8e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GCCCAGCGCTGGCGGAG 17
Db 2623 GCCCAGCGCTGGCGGAG 2607
|||||

RESULT 58
US-09-919-039-254/c
; Sequence 254, Application US/09919039
; Patent No. 6727066
; GENERAL INFORMATION:
; APPLICANT: Kaser, Matthew R.
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES
; FILE REFERENCE: PA-0035 US
; CURRENT APPLICATION NUMBER: US/09/919,039
; CURRENT FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: 60/222,113
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 401
; SOFTWARE: PERL Program
; SEQ ID NO 254
; LENGTH: 3814
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6727066 1081605.3
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 3792, 3796-3801
; OTHER INFORMATION: a, t, c, g, or other
; US-09-919-039-254

Query Match      81.2%; Score 13.8; DB 3; Length 3814;
Best Local Similarity 88.2%; Pred. No. 8.8e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GCCCAGCGCTGGCGGAG 17
Db 598 GCCCAGCGCTGGCGGAG 582
|||||

RESULT 59
US-09-252-991A-930/c
; Sequence 930, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 930
; LENGTH: 4563
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-930

Query Match      81.2%; Score 13.8; DB 3; Length 4563;
Best Local Similarity 88.2%; Pred. No. 8.8e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GCCCAGCGCTGGCGGAG 17
Db 3035 GCCCAGCGCTGGCGGTG 3019
|||||
```



```
RESULT 60
US-09-479-122-21
; Sequence 21, Application US/09479122
; Patent No. 6410266
; GENERAL INFORMATION:
; APPLICANT: HARRINGTON, JOHN J.
; APPLICANT: SHERF, BRUCE
; APPLICANT: RUNDLETT, STEPHEN
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; TITLE OF INVENTION: ENDOGENOUS GENES
; FILE REFERENCE: 0221-0003C
; CURRENT APPLICATION NUMBER: US/09/479,122
; CURRENT FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: 09/276,820
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 09/159,643
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 08/941,223
; PRIOR FILING DATE: 1997-09-26
; PRIOR APPLICATION NUMBER: 09/263,814
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/253,022
; PRIOR FILING DATE: 1999-02-19
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 5382
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (890)
; OTHER INFORMATION: a, c, t, g, other or unknown
; NAME/KEY: modified_base
; LOCATION: (1042)
; OTHER INFORMATION: a, c, t, g, other or unknown
; US-09-479-122-21

Query Match      81.2%; Score 13.8; DB 3; Length 5382;
Best Local Similarity 88.2%; Pred. No. 8.8e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1  GCCCAGCTTGCGCGAG 17
      ||||| ||||| |||||
Db      1501 GCCCGCGCATGCCGAG 1517

RESULT 61
US-09-484-997-21
; Sequence 21, Application US/09484997
; Patent No. 6524818
; GENERAL INFORMATION:
; APPLICANT: HARRINGTON, JOHN J.
; APPLICANT: SHERF, BRUCE
; APPLICANT: RUNDLETT, STEPHEN
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; TITLE OF INVENTION: ENDOGENOUS GENES
; FILE REFERENCE: 0221-0003U
; CURRENT APPLICATION NUMBER: US/09/484,997
; CURRENT FILING DATE: 2000-01-18
; PRIOR APPLICATION NUMBER: 09/276,820
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 09/159,643
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 08/941,223
; PRIOR FILING DATE: 1997-09-26
; PRIOR APPLICATION NUMBER: 09/263,814
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/253,022
; PRIOR FILING DATE: 1999-02-19
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21

Query Match      81.2%; Score 13.8; DB 3; Length 5382;
Best Local Similarity 88.2%; Pred. No. 8.8e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1  GCCCAGCTTGCGCGAG 17
      ||||| ||||| |||||
Db      1501 GCCCGCGCATGCCGAG 1517

RESULT 62
US-09-481-355-21
; Sequence 21, Application US/09481355
; Patent No. 6524824
; GENERAL INFORMATION:
; APPLICANT: HARRINGTON, JOHN J.
; APPLICANT: SHERF, BRUCE
; APPLICANT: RUNDLETT, STEPHEN
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; TITLE OF INVENTION: ENDOGENOUS GENES
; FILE REFERENCE: 0221-0003F
; CURRENT APPLICATION NUMBER: US/09/481,355
; CURRENT FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 09/276,820
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 09/159,643
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 08/941,223
; PRIOR FILING DATE: 1997-09-26
; PRIOR APPLICATION NUMBER: 09/263,814
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/253,022
; PRIOR FILING DATE: 1999-02-19
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 5382
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (890)
; OTHER INFORMATION: a, c, t, g, other or unknown
; NAME/KEY: modified_base
; LOCATION: (1042)
; OTHER INFORMATION: a, c, t, g, other or unknown
; US-09-481-355-21

Query Match      81.2%; Score 13.8; DB 3; Length 5382;
Best Local Similarity 88.2%; Pred. No. 8.8e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1  GCCCAGCTTGCGCGAG 17
      ||||| ||||| |||||
Db      1501 GCCCGCGCATGCCGAG 1517

RESULT 63
US-09-481-282-21
; Sequence 21, Application US/09481282
; Patent No. 6541221
; GENERAL INFORMATION:
```

```
; LENGTH: 5382
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (890)
; OTHER INFORMATION: a, c, t, g, other or unknown
; NAME/KEY: modified_base
; LOCATION: (1042)
; OTHER INFORMATION: a, c, t, g, other or unknown
; US-09-484-997-21

Query Match      81.2%; Score 13.8; DB 3; Length 5382;
Best Local Similarity 88.2%; Pred. No. 8.8e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1  GCCCAGCTTGCGCGAG 17
      ||||| ||||| |||||
Db      1501 GCCCGCGCATGCCGAG 1517

RESULT 62
US-09-481-355-21
; Sequence 21, Application US/09481355
; Patent No. 6524824
; GENERAL INFORMATION:
; APPLICANT: HARRINGTON, JOHN J.
; APPLICANT: SHERF, BRUCE
; APPLICANT: RUNDLETT, STEPHEN
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; TITLE OF INVENTION: ENDOGENOUS GENES
; FILE REFERENCE: 0221-0003F
; CURRENT APPLICATION NUMBER: US/09/481,355
; CURRENT FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 09/276,820
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 09/159,643
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 08/941,223
; PRIOR FILING DATE: 1997-09-26
; PRIOR APPLICATION NUMBER: 09/263,814
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/253,022
; PRIOR FILING DATE: 1999-02-19
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 5382
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (890)
; OTHER INFORMATION: a, c, t, g, other or unknown
; NAME/KEY: modified_base
; LOCATION: (1042)
; OTHER INFORMATION: a, c, t, g, other or unknown
; US-09-481-355-21

Query Match      81.2%; Score 13.8; DB 3; Length 5382;
Best Local Similarity 88.2%; Pred. No. 8.8e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1  GCCCAGCTTGCGCGAG 17
      ||||| ||||| |||||
Db      1501 GCCCGCGCATGCCGAG 1517

RESULT 63
US-09-481-282-21
; Sequence 21, Application US/09481282
; Patent No. 6541221
; GENERAL INFORMATION:
```

```

;
;
; NAME/KEY: modified_base
; FEATURE:
;
; LOCATION: (890)
;
; TITLE OF INVENTION: ENDOGENOUS GENES
;
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; APPLICANT: RUNDLETT, STEPHEN

```

```
FILE REFERENCE: 0221-0003B
; CURRENT APPLICATION NUMBER: US/09/479,123
; CURRENT FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: 09/263,814
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/253,022
; PRIOR FILING DATE: 1999-02-19
; PRIOR APPLICATION NUMBER: 09/159,643
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 08/941,223
; PRIOR FILING DATE: 1997-09-26
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 5382
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (890)
; OTHER INFORMATION: a, c, t, g, other or unknown
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1042)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-09-479-123-21

Query Match      81.2%; Score 13.8; DB 3; Length 5382;
Best Local Similarity 88.2%; Pred. No. 8.8e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 GCCCAGCTTGCCGAG 17
      ||||| ||||| ||||| |||||
Db      1501 GCCCGCGCATGCCGAG 1517

RESULT 67
US-09-484-317A-21
; Sequence 21, Application US/09484317A
; Patent No. 6740503
; GENERAL INFORMATION:
; APPLICANT: HARRINGTON, JOHN J.
; APPLICANT: SHERF, BRUCE
; APPLICANT: RUNDLETT, STEPHEN
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; FILE REFERENCE: ATX-007CP4DV11
; CURRENT APPLICATION NUMBER: US/09/484,317A
; CURRENT FILING DATE: 2000-01-18
; PRIOR APPLICATION NUMBER: 09/276,820
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 09/263,814
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/253,022
; PRIOR FILING DATE: 1999-02-19
; PRIOR APPLICATION NUMBER: 09/159,643
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 08/941,223
; PRIOR FILING DATE: 1997-09-26
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 5382
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (890)
; OTHER INFORMATION: a, c, t, g, other or unknown
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1042)
; OTHER INFORMATION: a, c, t, g, other or unknown
```

```
US-09-484-317A-21

Query Match      81.2%; Score 13.8; DB 3; Length 5382;
Best Local Similarity 88.2%; Pred. No. 8.8e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 GCCCAGCTTGCCGAG 17
      ||||| ||||| ||||| |||||
Db      1501 GCCCGCGCATGCCGAG 1517

RESULT 68
US-09-276-820A-21
; Sequence 21, Application US/09276820A
; Patent No. 6897066
; GENERAL INFORMATION:
; APPLICANT: HARRINGTON, JOHN J.
; APPLICANT: SHERF, BRUCE
; APPLICANT: RUNDLETT, STEPHEN
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; FILE REFERENCE: 0221-0003US
; CURRENT APPLICATION NUMBER: US/09/276,820A
; CURRENT FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 09/159,643
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 08/941,223
; PRIOR FILING DATE: 1997-09-26
; PRIOR APPLICATION NUMBER: 09/263,814
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/253,022
; PRIOR FILING DATE: 1999-02-19
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 5382
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (890)
; OTHER INFORMATION: a, c, t, g, other or unknown
; NAME/KEY: modified_base
; LOCATION: (1042)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-09-276-820A-21

Query Match      81.2%; Score 13.8; DB 3; Length 5382;
Best Local Similarity 88.2%; Pred. No. 8.8e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 GCCCAGCTTGCCGAG 17
      ||||| ||||| ||||| |||||
Db      1501 GCCCGCGCATGCCGAG 1517

RESULT 69
US-09-760-897A-21
; Sequence 21, Application US/09760897A
; Patent No. 7033782
; GENERAL INFORMATION:
; APPLICANT: HARRINGTON, JOHN J.
; APPLICANT: SHERF, BRUCE
; APPLICANT: RUNDLETT, STEPHEN
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; FILE REFERENCE: ATX7CP4D15CNRCE
; CURRENT APPLICATION NUMBER: US/09/760,897A
; CURRENT FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/515,124
; PRIOR FILING DATE: 2000-02-27
; PRIOR APPLICATION NUMBER: 09/276,820
; PRIOR FILING DATE: 1999-03-26
```

```
; PRIOR APPLICATION NUMBER: 09/263,814
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/253,022
; PRIOR FILING DATE: 1999-02-19
; PRIOR APPLICATION NUMBER: 09/159,643
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 08/941,223
; PRIOR FILING DATE: 1997-09-26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 5382
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (890)
; OTHER INFORMATION: a, c, t, g, other or unknown
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1042)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-09-760-897A-21

Query Match      81.2%; Score 13.8; DB 5; Length 5382;
Best Local Similarity 88.2%; Pred. No. 8.8e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 GCCCAGCGTTGGCCGAG 17
Db      1501 GCCCGCGCATGGCCGAG 1517

RESULT 70
US-08-487-283A-4
; Sequence 4, Application US/08487283A
; Patent No. 6355245
; GENERAL INFORMATION:
; APPLICANT: Evans, Mark J.
; APPLICANT: Matis, Louis A.
; APPLICANT: Mueller, Eileen Elliott
; APPLICANT: Nye, Steven H.
; APPLICANT: Rollins, Scott
; APPLICANT: Rother, Russell P.
; APPLICANT: Springhorn, Jeremy P.
; APPLICANT: Squinto, Stephen P.
; APPLICANT: Thomas, Thomas C.
; APPLICANT: Wilkins, James A.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE TREATMENT
; OF INFLAMMATORY DISEASES
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seth A. Fidel
; STREET: 25 Science Park (Alexion)
; CITY: New Haven
; STATE: Connecticut
; COUNTRY: USA
; ZIP: 06511
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.4Mb storage
; COMPUTER: Macintosh Centris 610
; OPERATING SYSTEM: System 7
; SOFTWARE: WordPerfect 3.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,283A
; FILING DATE: June 7, 1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/236,208
; FILING DATE: 02-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Seth A. Fidel.
; REGISTRATION NUMBER: 38,449
; REFERENCE/DOCKET NUMBER: ALX-152.1 CIP

; PRIOR APPLICATION INFORMATION:
; TELEPHONE: (203)776-1790
; TELEFAX: (203)772-3655
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8540 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Double
; TOPOLOGY: Circular
; MOLECULE TYPE: Other nucleic acid
; DESCRIPTION: Apex-3P Eukaryotic
; DESCRIPTION: Expression Vector
US-08-487-283A-4

Query Match      81.2%; Score 13.8; DB 3; Length 8540;
Best Local Similarity 88.2%; Pred. No. 8.8e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 GCCCAGCGTTGGCCGAG 17
Db      3112 GCCCGCGCATGGCCGAG 3128

RESULT 71
PCT-US96-05611A-12
; Sequence 12, Application PC/TUS9605611A
; GENERAL INFORMATION:
; APPLICANT: Mueller, John P.
; APPLICANT: Lenardo, Michael J.
; APPLICANT: McFarland, Henry F.
; APPLICANT: Matis, Louis A.
; APPLICANT: Mueller, Eileen Elliott
; APPLICANT: Nye, Steven H.
; APPLICANT: Pelfrey, Clara M.
; APPLICANT: Squinto, Stephen P.
; APPLICANT: Wilkins, James A.
; TITLE OF INVENTION: Modified Myelin Protein Molecules
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Maurice M. Klee
; STREET: 1951 Burr Street
; CITY: Fairfield
; STATE: Connecticut
; COUNTRY: USA
; ZIP: 06430
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 0.8 Mb storage
; COMPUTER: Macintosh Centris 610
; OPERATING SYSTEM: System 7
; SOFTWARE: Microsoft Word 6.0.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/05611A
; FILING DATE: 02-MAY-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/431,644
; FILING DATE: May 2, 1995
; APPLICATION NUMBER: 08/431,648
; FILING DATE: May 2, 1995
; APPLICATION NUMBER: 08/482,114
; FILING DATE: June 7, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Klee, Maurice M.
; REGISTRATION NUMBER: 30,399
; REFERENCE/DOCKET NUMBER: ALX-129
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (203) 255 1400
; TELEFAX: (203) 254 1101
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8540 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Double
```

```
; TOPOLOGY: Circular
; MOLECULE TYPE: Other nucleic acid
; DESCRIPTION: Apex-3P Eukaryotic
; DESCRIPTION: Expression Vector
PCT-US96-05611A-12

Query Match      81.2%; Score 13.8; DB 7; Length 8540;
Best Local Similarity 88.2%; Pred. No. 8.8e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGCCCGAG 17
    ||||| ||| |||||
Db 3112 GCCCGCATGCCCGAG 3128

RESULT 72
US-09-479-122-22
; Sequence 22, Application US/09479122
; Patent No. 6410266
; GENERAL INFORMATION:
; APPLICANT: HARRINGTON, JOHN J.
; APPLICANT: SHERF, BRUCE
; APPLICANT: RUNDLETT, STEPHEN
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; FILE REFERENCE: 0221-0003C
; CURRENT APPLICATION NUMBER: US/09/479,122
; CURRENT FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: 09/276,820
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 09/159,643
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 08/941,223
; PRIOR FILING DATE: 1997-09-26
; PRIOR APPLICATION NUMBER: 09/263,814
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/253,022
; PRIOR FILING DATE: 1999-02-19
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 9737
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8347)
; OTHER INFORMATION: a, c, t, g, other or unknown
; NAME/KEY: modified_base
; LOCATION: (8499)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-09-479-122-22

Query Match      81.2%; Score 13.8; DB 3; Length 9737;
Best Local Similarity 88.2%; Pred. No. 8.8e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGCCCGAG 17
    ||||| ||| |||||
Db 8958 GCCCGCATGCCCGAG 8974

RESULT 73
US-09-479-122-23
; Sequence 23, Application US/09479122
; Patent No. 6410266
; GENERAL INFORMATION:
; APPLICANT: HARRINGTON, JOHN J.
; APPLICANT: SHERF, BRUCE
; APPLICANT: RUNDLETT, STEPHEN
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; FILE REFERENCE: 0221-0003C
```

```
; CURRENT APPLICATION NUMBER: US/09/479,122
; CURRENT FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: 09/276,820
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 09/159,643
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 08/941,223
; PRIOR FILING DATE: 1997-09-26
; PRIOR APPLICATION NUMBER: 09/263,814
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/253,022
; PRIOR FILING DATE: 1999-02-19
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 23
; LENGTH: 9737
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8347)
; OTHER INFORMATION: a, c, t, g, other or unknown
; NAME/KEY: modified_base
; LOCATION: (8499)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-09-479-122-23

Query Match      81.2%; Score 13.8; DB 3; Length 9737;
Best Local Similarity 88.2%; Pred. No. 8.8e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGCCCGAG 17
    ||||| ||| |||||
Db 8958 GCCCGCATGCCCGAG 8974

RESULT 74
US-09-479-122-28
; Sequence 28, Application US/09479122
; Patent No. 6410266
; GENERAL INFORMATION:
; APPLICANT: HARRINGTON, JOHN J.
; APPLICANT: SHERF, BRUCE
; APPLICANT: RUNDLETT, STEPHEN
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; FILE REFERENCE: 0221-0003C
; CURRENT APPLICATION NUMBER: US/09/479,122
; CURRENT FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: 09/276,820
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 09/159,643
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 08/941,223
; PRIOR FILING DATE: 1997-09-26
; PRIOR APPLICATION NUMBER: 09/263,814
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/253,022
; PRIOR FILING DATE: 1999-02-19
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
; LENGTH: 9737
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8347)
; OTHER INFORMATION: a, c, t, g, other or unknown
; NAME/KEY: modified_base
; LOCATION: (8499)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-09-479-122-28
```

```
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 08/941,223
; PRIOR FILING DATE: 1997-09-26
; PRIOR APPLICATION NUMBER: 09/263,814
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/253,022
; PRIOR FILING DATE: 1999-02-19
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 23
; LENGTH: 9737
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8347)
; OTHER INFORMATION: a, c, t, g, other or unknown
; NAME/KEY: modified_base
; LOCATION: (8499)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-09-484-997-23

Query Match      81.2%; Score 13.8; DB 3; Length 9737;
Best Local Similarity 88.2%; Pred. No. 8.8e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 GCCCAGCTTGCCGAG 17
      ||||| ||| |||||
Db      8958 GCCCGCATGCCGAG 8974

RESULT 75
US-09-484-997-22
; Sequence 22, Application US/09484997
; Patent No. 6524818
; GENERAL INFORMATION:
; APPLICANT: HARRINGTON, JOHN J.
; APPLICANT: SHERF, BRUCE
; APPLICANT: RUNDLETT, STEPHEN
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; FILE REFERENCE: 0221-0003J
; CURRENT APPLICATION NUMBER: US/09/484,997
; PRIOR FILING DATE: 2000-01-18
; PRIOR APPLICATION NUMBER: 09/276,820
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 09/159,643
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 08/941,223
; PRIOR FILING DATE: 1997-09-26
; PRIOR APPLICATION NUMBER: 09/263,814
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/253,022
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 9737
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8347)
; OTHER INFORMATION: a, c, t, g, other or unknown
; NAME/KEY: modified_base
; LOCATION: (8499)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-09-484-997-22

Query Match      81.2%; Score 13.8; DB 3; Length 9737;
Best Local Similarity 88.2%; Pred. No. 8.8e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 GCCCAGCTTGCCGAG 17
      ||||| ||| |||||
Db      8958 GCCCGCATGCCGAG 8974

RESULT 76
US-09-484-997-23
; Sequence 23, Application US/09484997
; Patent No. 6524818
; GENERAL INFORMATION:
; APPLICANT: HARRINGTON, JOHN J.
; APPLICANT: SHERF, BRUCE
; APPLICANT: RUNDLETT, STEPHEN
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; FILE REFERENCE: 0221-0003J
; CURRENT APPLICATION NUMBER: US/09/484,997
; PRIOR FILING DATE: 2000-01-18
; PRIOR APPLICATION NUMBER: 09/276,820
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 09/159,643
```

```
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 08/941,223
; PRIOR FILING DATE: 1997-09-26
; PRIOR APPLICATION NUMBER: 09/263,814
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/253,022
; PRIOR FILING DATE: 1999-02-19
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 23
; LENGTH: 9737
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8347)
; OTHER INFORMATION: a, c, t, g, other or unknown
; NAME/KEY: modified_base
; LOCATION: (8499)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-09-484-997-23

Query Match      81.2%; Score 13.8; DB 3; Length 9737;
Best Local Similarity 88.2%; Pred. No. 8.8e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 GCCCAGCTTGCCGAG 17
      ||||| ||| |||||
Db      8958 GCCCGCATGCCGAG 8974

RESULT 77
US-09-484-997-28
; Sequence 28, Application US/09484997
; Patent No. 6524818
; GENERAL INFORMATION:
; APPLICANT: HARRINGTON, JOHN J.
; APPLICANT: SHERF, BRUCE
; APPLICANT: RUNDLETT, STEPHEN
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; FILE REFERENCE: 0221-0003J
; CURRENT APPLICATION NUMBER: US/09/484,997
; CURRENT FILING DATE: 2000-01-18
; PRIOR APPLICATION NUMBER: 09/276,820
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 09/159,643
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 08/941,223
; PRIOR FILING DATE: 1997-09-26
; PRIOR APPLICATION NUMBER: 09/263,814
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/253,022
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
; LENGTH: 9737
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8347)
; OTHER INFORMATION: a, c, t, g, other or unknown
; NAME/KEY: modified_base
; LOCATION: (8499)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-09-484-997-28

Query Match      81.2%; Score 13.8; DB 3; Length 9737;
Best Local Similarity 88.2%; Pred. No. 8.8e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```



```
RESULT 81
US-09-481-282-22
; Sequence 22, Application US/09481282
; Patent No. 6541221
; GENERAL INFORMATION:
; APPLICANT: HARRINGTON, JOHN J.
; APPLICANT: SHERF, BRUCE
; APPLICANT: RUNDLETT, STEPHEN
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; FILE REFERENCE: 0221-0003US
; CURRENT APPLICATION NUMBER: US/09/481,282
; CURRENT FILING DATE: 2000-01-11
; PRIOR FILING DATE: 1998-09-24
; PRIOR FILING DATE: 1998-09-24
; PRIOR FILING DATE: 1997-09-26
; PRIOR FILING DATE: 1997-09-26
; PRIOR FILING DATE: 1999-03-08
; PRIOR FILING DATE: 1999-03-08
; PRIOR FILING DATE: 1999-02-19
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 9737
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: modified_base
; LOCATION: (8347)
; OTHER INFORMATION: a, c, t, g, other or unknown
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8499)
; OTHER INFORMATION: a, c, t, g, other or unknown
; US-09-481-282-22
;
; Query Match      81.2%; Score 13.8; DB 3; Length 9737;
; Best Local Similarity 88.2%; Pred. No. 8.8e+02;
; Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
;
Qy 1 GCCCAGCTTGGCGGAG 17
    ||||| ||||| |||||
Db 8958 GCCCGCGCATGGCGGAG 8974

RESULT 82
US-09-481-282-23
; Sequence 23, Application US/09481282
; Patent No. 6541221
; GENERAL INFORMATION:
; APPLICANT: HARRINGTON, JOHN J.
; APPLICANT: SHERF, BRUCE
; APPLICANT: RUNDLETT, STEPHEN
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; FILE REFERENCE: 0221-0003US
; CURRENT APPLICATION NUMBER: US/09/481,282
; CURRENT FILING DATE: 2000-01-11
; PRIOR FILING DATE: 1998-09-24
; PRIOR FILING DATE: 1998-09-24
; PRIOR FILING DATE: 1997-09-26
; PRIOR FILING DATE: 1997-09-26
; PRIOR FILING DATE: 1999-03-08
; PRIOR FILING DATE: 1999-03-08
; PRIOR FILING DATE: 1999-02-19
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 23
; LENGTH: 9737
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: modified_base
; LOCATION: (8347)
; OTHER INFORMATION: a, c, t, g, other or unknown
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8499)
; OTHER INFORMATION: a, c, t, g, other or unknown
; US-09-481-282-23
;
; Query Match      81.2%; Score 13.8; DB 3; Length 9737;
; Best Local Similarity 88.2%; Pred. No. 8.8e+02;
; Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
;
Qy 1 GCCCAGCTTGGCGGAG 17
    ||||| ||||| |||||
Db 8958 GCCCGCGCATGGCGGAG 8974

RESULT 83
US-09-481-282-28
; Sequence 28, Application US/09481282
; Patent No. 6541221
; GENERAL INFORMATION:
; APPLICANT: HARRINGTON, JOHN J.
; APPLICANT: SHERF, BRUCE
; APPLICANT: RUNDLETT, STEPHEN
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; FILE REFERENCE: 0221-0003US
; CURRENT APPLICATION NUMBER: US/09/481,282
; CURRENT FILING DATE: 2000-01-11
; PRIOR FILING DATE: 1998-09-24
; PRIOR FILING DATE: 1998-09-24
; PRIOR FILING DATE: 1997-09-26
; PRIOR FILING DATE: 1997-09-26
; PRIOR FILING DATE: 1999-03-08
; PRIOR FILING DATE: 1999-03-08
; PRIOR FILING DATE: 1999-02-19
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
; LENGTH: 9737
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: modified_base
; LOCATION: (8347)
; OTHER INFORMATION: a, c, t, g, other or unknown
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8499)
; OTHER INFORMATION: a, c, t, g, other or unknown
; US-09-481-282-28
;
; Query Match      81.2%; Score 13.8; DB 3; Length 9737;
; Best Local Similarity 88.2%; Pred. No. 8.8e+02;
; Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
;
Qy 1 GCCCAGCTTGGCGGAG 17
    ||||| ||||| |||||
Db 8958 GCCCGCGCATGGCGGAG 8974

RESULT 84
US-09-455-659A-22
; Sequence 22, Application US/09455659A
; Patent No. 6602686
; GENERAL INFORMATION:
; APPLICANT: HARRINGTON, JOHN J.
; APPLICANT: SHERF, BRUCE
; APPLICANT: RUNDLETT, STEPHEN
```



```
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; FILE REFERENCE: 0221-0003A
; CURRENT APPLICATION NUMBER: US/09/455,659A
; CURRENT FILING DATE: 1999-12-07
; PRIOR APPLICATION NUMBER: 09/276,820
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 09/263,814
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/253,022
; PRIOR FILING DATE: 1999-02-19
; PRIOR APPLICATION NUMBER: 09/159,643
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 08/941,223
; PRIOR FILING DATE: 1997-09-26
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 9737
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8347)
; OTHER INFORMATION: a, c, t, g, other or unknown
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; FILE REFERENCE: 0221-0003A
; CURRENT APPLICATION NUMBER: US/09/455,659A
; CURRENT FILING DATE: 1999-12-07
; PRIOR APPLICATION NUMBER: 09/276,820
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 09/263,814
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/253,022
; PRIOR FILING DATE: 1999-02-19
; PRIOR APPLICATION NUMBER: 09/159,643
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 08/941,223
; PRIOR FILING DATE: 1997-09-26
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 23
; LENGTH: 9737
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8347)
; OTHER INFORMATION: a, c, t, g, other or unknown
```

```
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8499)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-09-455-659A-23

Query Match      81.2%; Score 13.8; DB 3; Length 9737;
Best Local Similarity 88.2%; Pred. No. 8.8e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1  GCCCAGCTTGCCGAG 17
      ||||| ||||| |||||
Db      8958  GCCCGCATGCCGAG 8974

RESULT 86
US-09-455-659A-28
; Sequence 28, Application US/09455659A
; Patent No. 6602686
; GENERAL INFORMATION:
; APPLICANT: HARRINGTON, JOHN J.
; APPLICANT: SHERF, BRUCE
; APPLICANT: RUNDLETT, STEPHEN
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; TITLE OF INVENTION: ENDOGENOUS GENES
; FILE REFERENCE: 0221-0003A
; CURRENT APPLICATION NUMBER: US/09/455,659A
; CURRENT FILING DATE: 1999-12-07
; PRIOR APPLICATION NUMBER: 09/276,820
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 09/263,814
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/253,022
; PRIOR FILING DATE: 1999-02-19
; PRIOR APPLICATION NUMBER: 09/159,643
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 08/941,223
; PRIOR FILING DATE: 1997-09-26
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
; LENGTH: 9737
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8347)
; OTHER INFORMATION: a, c, t, g, other or unknown
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; FILE REFERENCE: 0221-0003A
; CURRENT APPLICATION NUMBER: US/09/455,659A
; CURRENT FILING DATE: 1999-12-07
; PRIOR APPLICATION NUMBER: 09/276,820
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 09/263,814
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/253,022
; PRIOR FILING DATE: 1999-02-19
; PRIOR APPLICATION NUMBER: 09/159,643
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 08/941,223
; PRIOR FILING DATE: 1997-09-26
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
; LENGTH: 9737
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8347)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-09-455-659A-28

Query Match      81.2%; Score 13.8; DB 3; Length 9737;
Best Local Similarity 88.2%; Pred. No. 8.8e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1  GCCCAGCTTGCCGAG 17
      ||||| ||||| |||||
Db      8958  GCCCGCATGCCGAG 8974

RESULT 87
US-09-484-996-22
; Sequence 22, Application US/09484996
; Patent No. 6623958
; GENERAL INFORMATION:
; APPLICANT: HARRINGTON, JOHN J.
; APPLICANT: SHERF, BRUCE
; APPLICANT: RUNDLETT, STEPHEN
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; TITLE OF INVENTION: ENDOGENOUS GENES
```

```
; FILE REFERENCE: 0221-0003H
; CURRENT APPLICATION NUMBER: US/09/484,996
; CURRENT FILING DATE: 2000-01-18
; PRIOR APPLICATION NUMBER: 09/276,820
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 09/263,814
; PRIOR FILING DATE: 1998-03-08
; PRIOR APPLICATION NUMBER: 09/253,022
; PRIOR FILING DATE: 1999-02-19
; PRIOR APPLICATION NUMBER: 09/159,643
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 08/941,223
; PRIOR FILING DATE: 1997-09-26
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 9737
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8347)
; OTHER INFORMATION: a, c, t, g, other or unknown
; NAME/KEY: modified_base
; LOCATION: (8499)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-09-484-996-22

Query Match      81.2%; Score 13.8; DB 3; Length 9737;
Best Local Similarity 88.2%; Pred. No. 8.8e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1  GCCCAGCTTGCCGCGAG 17
      ||||| ||||| ||||| |||||
Db      8958 GCCCGCGCATGCCGCGAG 8974

RESULT 89
US-09-484-996-28
; Sequence 28, Application US/09484996
; Patent No. 6623958
; GENERAL INFORMATION:
; APPLICANT: HARRINGTON, JOHN J.
; APPLICANT: SHERP, BRUCE
; APPLICANT: RUNDLETT, STEPHEN
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; TITLE OF INVENTION: ENDOGENOUS GENES
; FILE REFERENCE: 0221-0003H
; CURRENT APPLICATION NUMBER: US/09/484,996
; CURRENT FILING DATE: 2000-01-18
; PRIOR APPLICATION NUMBER: 09/276,820
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 09/263,814
; PRIOR FILING DATE: 1998-03-08
; PRIOR APPLICATION NUMBER: 09/253,022
; PRIOR FILING DATE: 1999-02-19
; PRIOR APPLICATION NUMBER: 09/159,643
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 08/941,223
; PRIOR FILING DATE: 1997-09-26
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
; LENGTH: 9737
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8347)
; OTHER INFORMATION: a, c, t, g, other or unknown
; NAME/KEY: modified_base
; LOCATION: (8499)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-09-484-996-28

Query Match      81.2%; Score 13.8; DB 3; Length 9737;
Best Local Similarity 88.2%; Pred. No. 8.8e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1  GCCCAGCTTGCCGCGAG 17
      ||||| ||||| ||||| |||||
Db      8958 GCCCGCGCATGCCGCGAG 8974

RESULT 90
US-09-479-123-22
; Sequence 22, Application US/09479123
; Patent No. 6670185
; GENERAL INFORMATION:
; APPLICANT: HARRINGTON, JOHN J.
; APPLICANT: SHERP, BRUCE
; APPLICANT: RUNDLETT, STEPHEN
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; TITLE OF INVENTION: ENDOGENOUS GENES
; FILE REFERENCE: 0221-0003B
; CURRENT APPLICATION NUMBER: US/09/479,123
; CURRENT FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: 09/263,814
; PRIOR FILING DATE: 1999-03-08
```

```

; PRIOR APPLICATION NUMBER: 09/253,022
; PRIOR FILING DATE: 1999-02-19
; PRIOR APPLICATION NUMBER: 09/159,643
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 08/941,223
; PRIOR FILING DATE: 1997-09-26
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 9737
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8347)
; OTHER INFORMATION: a, c, t, g, other or unknown
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8499)
; OTHER INFORMATION: a, c, t, g, other or unknown
; US-09-479-123-22

Query Match      81.2%; Score 13.8; DB 3; Length 9737;
Best Local Similarity 88.2%; Pred. No. 8.8e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1  GCCCAGCTTGCCGAG 17
Db      8958  GCCCGCATGCCGAG 8974

RESULT 91
US-09-479-123-23
; Sequence 23, Application US/09479123
; Patent No. 6670185
; GENERAL INFORMATION:
; APPLICANT: HARRINGTON, JOHN J.
; APPLICANT: SHERP, BRUCE
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; FILE REFERENCE: 0221-0003B
; CURRENT APPLICATION NUMBER: US/09/479,123
; CURRENT FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: 09/263,814
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/253,022
; PRIOR FILING DATE: 1999-02-19
; PRIOR APPLICATION NUMBER: 09/159,643
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 08/941,223
; PRIOR FILING DATE: 1997-09-26
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 23
; LENGTH: 9737
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8347)
; OTHER INFORMATION: a, c, t, g, other or unknown
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8499)
; OTHER INFORMATION: a, c, t, g, other or unknown
; US-09-479-123-23

Query Match      81.2%; Score 13.8; DB 3; Length 9737;
Best Local Similarity 88.2%; Pred. No. 8.8e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1  GCCCAGCTTGCCGAG 17
Db      8958  GCCCGCATGCCGAG 8974

RESULT 92
US-09-479-123-28
; Sequence 28, Application US/09479123
; Patent No. 6670185
; GENERAL INFORMATION:
; APPLICANT: HARRINGTON, JOHN J.
; APPLICANT: SHERP, BRUCE
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; FILE REFERENCE: 0221-0003B
; CURRENT APPLICATION NUMBER: US/09/479,123
; CURRENT FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: 09/263,814
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/253,022
; PRIOR FILING DATE: 1999-02-19
; PRIOR APPLICATION NUMBER: 09/159,643
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 08/941,223
; PRIOR FILING DATE: 1997-09-26
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
; LENGTH: 9737
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8347)
; OTHER INFORMATION: a, c, t, g, other or unknown
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8499)
; OTHER INFORMATION: a, c, t, g, other or unknown
; US-09-479-123-28

Query Match      81.2%; Score 13.8; DB 3; Length 9737;
Best Local Similarity 88.2%; Pred. No. 8.8e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1  GCCCAGCTTGCCGAG 17
Db      8958  GCCCGCATGCCGAG 8974

RESULT 93
US-09-484-317A-22
; Sequence 22, Application US/09484317A
; Patent No. 6740503
; GENERAL INFORMATION:
; APPLICANT: HARRINGTON, JOHN J.
; APPLICANT: SHERP, BRUCE
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; FILE REFERENCE: ATX-007CP4DV11
; CURRENT APPLICATION NUMBER: US/09/484,317A
; CURRENT FILING DATE: 2000-01-18
; PRIOR APPLICATION NUMBER: 09/276,820
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 09/263,814
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/253,022
; PRIOR FILING DATE: 1999-02-19
; PRIOR APPLICATION NUMBER: 09/159,643
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 08/941,223
; PRIOR FILING DATE: 1997-09-26
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 23
; LENGTH: 9737
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8347)
; OTHER INFORMATION: a, c, t, g, other or unknown
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8499)
; OTHER INFORMATION: a, c, t, g, other or unknown
; US-09-479-123-23

Query Match      81.2%; Score 13.8; DB 3; Length 9737;
Best Local Similarity 88.2%; Pred. No. 8.8e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1  GCCCAGCTTGCCGAG 17
Db      8958  GCCCGCATGCCGAG 8974
```

```
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 9737
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8347)
; OTHER INFORMATION: a, c, t, g, other or unknown
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8499)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-09-484-317A-22

Query Match      81.2%; Score 13.8; DB 3; Length 9737;
Best Local Similarity 88.2%; Pred. No. 8.8e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 GCCCAGCTTGCCGCGAG 17
      ||||| ||||| ||||| |||||
Db      8958 GCCCGCGCATGCCGCGAG 8974

RESULT 94
US-09-484-317A-23
; Sequence 23, Application US/09484317A
; Patent No. 6740503
; GENERAL INFORMATION:
; APPLICANT: HARRINGTON, JOHN J.
; APPLICANT: SHERF, BRUCE
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; FILE REFERENCE: ATX-007CP4DV11
; CURRENT APPLICATION NUMBER: US/09/484,317A
; CURRENT FILING DATE: 2000-01-18
; PRIOR APPLICATION NUMBER: 09/276,820
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 09/263,814
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/253,022
; PRIOR FILING DATE: 1999-02-19
; PRIOR APPLICATION NUMBER: 09/159,643
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 08/941,223
; PRIOR FILING DATE: 1997-09-26
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 23
; LENGTH: 9737
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8347)
; OTHER INFORMATION: a, c, t, g, other or unknown
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8499)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-09-484-317A-23

Query Match      81.2%; Score 13.8; DB 3; Length 9737;
Best Local Similarity 88.2%; Pred. No. 8.8e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 GCCCAGCTTGCCGCGAG 17
      ||||| ||||| ||||| |||||
Db      8958 GCCCGCGCATGCCGCGAG 8974

RESULT 95
US-09-484-317A-28
; Sequence 28, Application US/09484317A
; Patent No. 6740503
; GENERAL INFORMATION:
; APPLICANT: HARRINGTON, JOHN J.
; APPLICANT: SHERF, BRUCE
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; FILE REFERENCE: ATX-007CP4DV11
; CURRENT APPLICATION NUMBER: US/09/484,317A
; CURRENT FILING DATE: 2000-01-18
; PRIOR APPLICATION NUMBER: 09/276,820
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 09/263,814
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/253,022
; PRIOR FILING DATE: 1999-02-19
; PRIOR APPLICATION NUMBER: 09/159,643
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 08/941,223
; PRIOR FILING DATE: 1997-09-26
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
; LENGTH: 9737
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8347)
; OTHER INFORMATION: a, c, t, g, other or unknown
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8499)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-09-484-317A-28

Query Match      81.2%; Score 13.8; DB 3; Length 9737;
Best Local Similarity 88.2%; Pred. No. 8.8e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 GCCCAGCTTGCCGCGAG 17
      ||||| ||||| ||||| |||||
Db      8958 GCCCGCGCATGCCGCGAG 8974

RESULT 96
US-09-276-820A-22
; Sequence 22, Application US/09276820A
; Patent No. 6897066
; GENERAL INFORMATION:
; APPLICANT: HARRINGTON, JOHN J.
; APPLICANT: SHERF, BRUCE
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; FILE REFERENCE: 0221-0003US
; CURRENT APPLICATION NUMBER: US/09/276,820A
; CURRENT FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 09/159,643
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 08/941,223
; PRIOR FILING DATE: 1997-09-26
; PRIOR APPLICATION NUMBER: 09/263,814
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/253,022
; PRIOR FILING DATE: 1999-02-19
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 9737
```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8347)
; OTHER INFORMATION: a, c, t, g, other or unknown
; NAME/KEY: modified_base
; LOCATION: (8499)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-09-276-820A-22

Query Match      81.2%; Score 13.8; DB 3; Length 9737;
Best Local Similarity 88.2%; Pred. No. 8.8e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCCGAG 17
    ||||| ||||| ||||| |||||
Db 8958 GCCCGCGCATGCCGAG 8974

RESULT 97
US-09-276-820A-23
; Sequence 23, Application US/09276820A
; Patent No. 6897066
; GENERAL INFORMATION:
; APPLICANT: HARRINGTON, JOHN J.
; APPLICANT: SHERF, BRUCE
; APPLICANT: RUNDLETT, STEPHEN
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; FILE REFERENCE: 0221-0003US
; CURRENT APPLICATION NUMBER: US/09/276,820A
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 09/159,643
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 08/941,223
; PRIOR FILING DATE: 1997-09-26
; PRIOR APPLICATION NUMBER: 09/263,814
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/253,022
; PRIOR FILING DATE: 1999-02-19
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 23
; LENGTH: 9737
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8347)
; OTHER INFORMATION: a, c, t, g, other or unknown
; NAME/KEY: modified_base
; LOCATION: (8499)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-09-276-820A-23

Query Match      81.2%; Score 13.8; DB 3; Length 9737;
Best Local Similarity 88.2%; Pred. No. 8.8e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCCGAG 17
    ||||| ||||| ||||| |||||
Db 8958 GCCCGCGCATGCCGAG 8974

RESULT 98
US-09-276-820A-28
; Sequence 28, Application US/09276820A
; Patent No. 6897066
; GENERAL INFORMATION:
; APPLICANT: HARRINGTON, JOHN J.
; APPLICANT: SHERF, BRUCE
; APPLICANT: RUNDLETT, STEPHEN
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; FILE REFERENCE: 0221-0003US
; CURRENT APPLICATION NUMBER: US/09/276,820A
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 09/159,643
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 08/941,223
; PRIOR FILING DATE: 1997-09-26
; PRIOR APPLICATION NUMBER: 09/263,814
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/253,022
; PRIOR FILING DATE: 1999-02-19
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
; LENGTH: 9737
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8347)
; OTHER INFORMATION: a, c, t, g, other or unknown
; NAME/KEY: modified_base
; LOCATION: (8499)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-09-276-820A-28

Query Match      81.2%; Score 13.8; DB 3; Length 9737;
Best Local Similarity 88.2%; Pred. No. 8.8e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCCGAG 17
    ||||| ||||| ||||| |||||
Db 8958 GCCCGCGCATGCCGAG 8974

RESULT 99
US-09-760-897A-22
; Sequence 22, Application US/09760897A
; Patent No. 7033782
; GENERAL INFORMATION:
; APPLICANT: HARRINGTON, JOHN J.
; APPLICANT: SHERF, BRUCE
; APPLICANT: RUNDLETT, STEPHEN
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; FILE REFERENCE: ATX7CP4D15CNRC
; CURRENT APPLICATION NUMBER: US/09/760,897A
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/515,124
; PRIOR FILING DATE: 2000-02-27
; PRIOR APPLICATION NUMBER: 09/276,820
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 09/263,814
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/253,022
; PRIOR FILING DATE: 1999-02-19
; PRIOR APPLICATION NUMBER: 09/159,643
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 08/941,223
; PRIOR FILING DATE: 1997-09-26
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 9737
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8347)
; OTHER INFORMATION: a, c, t, g, other or unknown
; NAME/KEY: modified_base
; LOCATION: (8499)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-09-760-897A-22
```

```
; NAME/KEY: modified_base
; LOCATION: (8499)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-09-760-897A-22

Query Match      81.2%; Score 13.8; DB 5; Length 9737;
Best Local Similarity 88.2%; Pred. NO. 8.8e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 GCCCAGCGCTTGGCCGAG 17
      ||||| ||||| |||||
Db      8958 GCCCGCGCATGGCCGAG 8974

RESULT 100
US-09-760-897A-23
; Sequence 23, Application US/09760897A
; Patent No. 7033782
; GENERAL INFORMATION:
; APPLICANT: HARRINGTON, JOHN J.
; APPLICANT: SHERF, BRUCE
; APPLICANT: RUNDLETT, STEPHEN
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; TITLE OF INVENTION: ENDOGENOUS GENES
; FILE REFERENCE: ATX7CP4D15CNRC
; CURRENT APPLICATION NUMBER: US/09760,897A
; CURRENT FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/515,124
; PRIOR FILING DATE: 2000-02-27
; PRIOR APPLICATION NUMBER: 09/276,820
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 09/263,814
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/253,022
; PRIOR FILING DATE: 1999-02-19
; PRIOR APPLICATION NUMBER: 09/159,643
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 08/941,223
; PRIOR FILING DATE: 1997-09-26
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 23
; LENGTH: 9737
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8347)
; OTHER INFORMATION: a, c, t, g, other or unknown
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8499)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-09-760-897A-23

Query Match      81.2%; Score 13.8; DB 5; Length 9737;
Best Local Similarity 88.2%; Pred. NO. 8.8e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 GCCCAGCGCTTGGCCGAG 17
      ||||| ||||| |||||
Db      8958 GCCCGCGCATGGCCGAG 8974
```

Search completed: June 10, 2006, 15:39:30  
Job time : 73.1083 secs

,  
.  
.  
.  
.

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2006, 15:26:46 ; Search time 651.1 Seconds  
(without alignments)  
320.826 Million cell updates/sec

Title: US-10-600-816-31

Perfect score: 17

Sequence: 1 gccacgcgttgccgag 17

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 18892170 seqs, 6143817638 residues

Total number of hits satisfying chosen parameters: 37784340

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

Database : Published Applications NA Main:\*

1: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
3: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*  
4: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*  
5: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*  
6: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*  
7: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*  
8: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10C\_PUBCOMB.seq.\*  
9: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10D\_PUBCOMB.seq.\*  
10: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10E\_PUBCOMB.seq.\*  
11: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10F\_PUBCOMB.seq.\*  
12: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10G\_PUBCOMB.seq.\*  
13: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US11A\_PUBCOMB.seq.\*  
14: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US11B\_PUBCOMB.seq.\*  
15: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US11C\_PUBCOMB.seq.\*  
16: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US11D\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	17	100.0	17	8	US-10-600-816-31
2	17	100.0	42	8	US-10-600-816-30
3	17	100.0	497	6	US-10-066-543-1937
C 4	17	100.0	552	6	US-10-066-543-1811
C 5	17	100.0	620	3	US-09-969-034-2222
C 6	17	100.0	634	3	US-09-969-034-3393
C 7	17	100.0	642	7	US-10-125-968-701
8	17	100.0	1071	9	US-10-112-615-134
9	17	100.0	1212	3	US-09-866-050A-249
10	17	100.0	1212	6	US-10-152-661-249
11	17	100.0	1228	7	US-10-313-542-223
12	17	100.0	1400	13	US-11-060-756-2418
13	17	100.0	1400	13	US-11-060-756-6690
14	17	100.0	1460	10	US-10-936-626-64
15	17	100.0	1460	10	US-10-938-061-64
16	17	100.0	1619	7	US-10-224-289-5
17	17	100.0	1619	9	US-10-935-190-43
18	17	100.0	1788	10	US-10-505-486-196
19	17	100.0	2297	9	US-10-775-920-13
20	17	100.0	2302	7	US-10-224-289-3
21	17	100.0	2302	8	US-10-240-425-405
22	17	100.0	2302	9	US-10-775-920-9
23	17	100.0	2302	10	US-10-510-507-2
24	17	100.0	2305	6	US-10-775-920-12
25	17	100.0	2316	6	US-10-176-847-59
26	17	100.0	2316	15	US-11-080-991-59
27	17	100.0	2446	9	US-10-775-920-11
28	17	100.0	2456	6	US-10-225-567A-453
29	17	100.0	2456	7	US-10-269-909-63
30	17	100.0	2456	7	US-10-269-909-64
31	17	100.0	2456	7	US-10-295-027-619
32	17	100.0	2456	8	US-10-600-816-2
33	17	100.0	2456	9	US-10-600-816-18
34	17	100.0	2456	9	US-10-775-920-10
35	17	100.0	2456	10	US-10-936-626-40
36	17	100.0	2456	10	US-10-938-061-40
37	17	100.0	2456	16	US-11-169-041-32
38	17	100.0	2593	7	US-10-264-049-834
39	17	100.0	4239	6	US-10-198-846-10424
C 40	15.4	90.6	1020	8	US-10-335-977-1640
C 41	15.4	90.6	1050	8	US-10-335-977-1641
42	15.4	90.6	1173	10	US-10-956-157-41
43	15.4	90.6	1173	10	US-10-956-157-5276
44	15.4	90.6	2306	8	US-10-398-221-3404
45	15.4	90.6	4183	3	US-09-757-716-2
46	15.4	90.6	4222	8	US-10-723-860-7430
47	15.4	90.6	63761	9	US-10-322-281-557
48	15.4	90.6	68732	8	US-10-322-281-560
49	15.4	90.6	495269	8	US-10-398-221-8
C 50	15.4	90.6	2256646	8	US-10-470-565-1
C 51	15.4	90.6	2944528	16	US-11-045-004-1
C 52	15.4	90.6	3011208	8	US-10-398-221-2058
C 53	15	88.2	337	9	US-10-357-930-57611
C 54	15	88.2	404	9	US-10-425-115-123425
C 55	15	88.2	3303	9	US-10-739-930-4894
C 56	14.4	84.7	225	10	US-10-956-157-193814
C 57	14.4	84.7	201	10	US-10-995-561-64090
C 58	14.4	84.7	344	3	US-09-803-719-429
C 59	14.4	84.7	344	10	US-10-779-543-13699
60	14.4	84.7	570	9	US-10-357-930-54041
C 61	14.4	84.7	653	4	US-09-925-065A-939940
C 62	14.4	84.7	653	5	US-09-925-065A-939940
C 63	14.4	84.7	834	4	US-09-925-065A-63251
C 64	14.4	84.7	834	5	US-09-925-065A-63251
C 65	14.4	84.7	834	12	US-10-301-480-164489
C 66	14.4	84.7	834	12	US-10-301-480-164489
C 67	14.4	84.7	1889	7	US-10-101-510-469
C 68	14.4	84.7	13608	3	US-09-861-846-3
C 69	14.4	84.7	13608	8	US-10-250-463-3
C 70	14.4	84.7	49312	8	US-10-294-934-485
C 71	14.4	84.7	49312	16	US-11-204-311-485
C 72	14.4	84.7	326002	10	US-10-461-862-56
73	14	82.4	25	8	US-10-713-956-525977
74	14	82.4	25	9	US-10-719-900-974704
C 75	14	82.4	201	10	US-10-995-561-63916
C 76	14	82.4	429	12	US-10-301-480-242988
C 77	14	82.4	429	12	US-10-301-480-856397
C 78	14	82.4	440	4	US-09-925-065A-148858
C 79	14	82.4	440	5	US-09-925-065A-148858
C 80	14	82.4	458	3	US-09-918-995-28686
81	14	82.4	465	9	US-10-674-134A-19328
82	14	82.4	518	5	US-09-925-065A-82285
83	14	82.4	518	5	US-09-925-065A-82285
84	14	82.4	518	12	US-10-301-480-183525
85	14	82.4	518	12	US-10-301-480-796934
C 86	14	82.4	614	4	US-09-925-065A-805681
C 87	14	82.4	614	5	US-09-925-065A-805681
88	14	82.4	683	4	US-09-925-065A-14
89	14	82.4	683	4	US-09-925-065A-15
90	14	82.4	683	5	US-09-925-065A-14

Sequence 196, App  
Sequence 13, Appl  
Sequence 3, Appl  
Sequence 405, App  
Sequence 9, Appl  
Sequence 2, Appl  
Sequence 12, Appl  
Sequence 59, Appl  
Sequence 11, Appl  
Sequence 453, App  
Sequence 63, Appl  
Sequence 64, Appl  
Sequence 619, App  
Sequence 2, Appl  
Sequence 10, Appl  
Sequence 18, Appl  
Sequence 40, Appl  
Sequence 40, Appl  
Sequence 32, Appl  
Sequence 834, App  
Sequence 10424, A  
Sequence 1640, Ap  
Sequence 1641, Ap  
Sequence 41, Appl  
Sequence 5276, Ap  
Sequence 3404, Ap  
Sequence 2, Appl  
Sequence 7430, Ap  
Sequence 557, App  
Sequence 560, App  
Sequence 8, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 2058, Ap  
Sequence 57611, A  
Sequence 123425,  
Sequence 193814,  
Sequence 64090, A  
Sequence 429, App  
Sequence 13699, A  
Sequence 54041, A  
Sequence 939940,  
Sequence 939940,  
Sequence 63251, A  
Sequence 63251, A  
Sequence 164489,  
Sequence 777898,  
Sequence 469, App  
Sequence 3, Appl  
Sequence 3, Appl  
Sequence 485, App  
Sequence 485, App  
Sequence 56, Appl  
Sequence 525977,  
Sequence 974704,  
Sequence 63916, A  
Sequence 242988,  
Sequence 856397,  
Sequence 148858,  
Sequence 148858,  
Sequence 28686, A  
Sequence 19328, A  
Sequence 82285, A  
Sequence 82285, A  
Sequence 183525,  
Sequence 786934,  
Sequence 805681,  
Sequence 805681,  
Sequence 14, Appl  
Sequence 15, Appl  
Sequence 14, Appl















c 967 13 76.5 616 12 US-10-301-480-523399 Sequence 525399,  
c 968 13 76.5 616 12 US-10-301-480-525400 Sequence 525400,  
c 969 13 76.5 616 12 US-10-301-480-1138805 Sequence 1138805,  
c 970 13 76.5 616 12 US-10-301-480-1138806 Sequence 1138806,  
c 971 13 76.5 616 12 US-10-301-480-1138807 Sequence 1138807,  
c 972 13 76.5 616 12 US-10-301-480-1138808 Sequence 1138808,  
c 973 13 76.5 616 12 US-10-301-480-1138809 Sequence 1138809,  
c 974 13 76.5 617 12 US-10-301-480-388662 Sequence 388662,  
c 975 13 76.5 617 12 US-10-301-480-1002071 Sequence 1002071,  
c 976 13 76.5 637 8 US-10-260-238-4465 Sequence 4465, Ap  
c 977 13 76.5 642 4 US-09-925-065A-700567 Sequence 700567,  
c 978 13 76.5 642 4 US-09-925-065A-700567 Sequence 700567,  
c 979 13 76.5 648 4 US-09-925-065A-509347 Sequence 509347,  
c 980 13 76.5 648 5 US-09-925-065A-509347 Sequence 509347,  
c 981 13 76.5 727 3 US-09-846-903-97 Sequence 97, Appl  
c 982 13 76.5 727 8 US-10-660-208-97 Sequence 97, Appl  
c 983 13 76.5 741 12 US-10-301-480-555185 Sequence 555185,  
c 984 13 76.5 741 12 US-10-301-480-1168594 Sequence 1168594,  
c 985 13 76.5 746 3 US-09-846-903-98 Sequence 98, Appl  
c 986 13 76.5 746 8 US-10-660-208-98 Sequence 98, Appl  
c 987 13 76.5 809 3 US-09-846-903-80 Sequence 80, Appl  
c 988 13 76.5 809 8 US-10-660-208-80 Sequence 80, Appl  
c 989 13 76.5 822 9 US-10-425-115-58342 Sequence 58342, A  
c 990 13 76.5 828 3 US-09-846-903-95 Sequence 95, Appl  
c 991 13 76.5 828 8 US-10-660-208-95 Sequence 95, Appl  
c 992 13 76.5 842 6 US-10-027-632-173732 Sequence 173732,  
c 993 13 76.5 842 7 US-10-027-632-173732 Sequence 173732,  
c 994 13 76.5 847 3 US-09-846-903-96 Sequence 96, Appl  
c 995 13 76.5 847 8 US-10-660-208-96 Sequence 96, Appl  
c 996 13 76.5 876 7 US-10-369-493-39867 Sequence 39867, A  
c 997 13 76.5 888 7 US-10-369-493-39485 Sequence 39485, A  
c 998 13 76.5 894 7 US-10-369-493-39117 Sequence 39117, A  
c 999 13 76.5 910 3 US-09-846-903-81 Sequence 81, Appl  
1000 13 76.5 910 8 US-10-660-208-81 Sequence 81, Appl

## ALIGNMENTS

RESULT 1  
US-10-600-816-31  
; Sequence 31, Application US/10600816  
; Publication No. US20040121362A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: IDENTIFICATION AND MODULATION OF A G-PROTEIN COUPLED RECEPTOR  
; TITLE OF INVENTION: (GPCR), RAI3, ASSOCIATED WITH CHRONIC OBSTRUCTIVE PULMONARY  
; TITLE OF INVENTION: DISEASE (COPD) AND NF-KB AND E-SELECTIN REGULATION  
; FILE REFERENCE: D0251 NP  
; CURRENT APPLICATION NUMBER: US/10/600,816  
; CURRENT FILING DATE: 2003-06-20  
; PRIOR APPLICATION NUMBER: U.S. 60/390,850  
; PRIOR FILING DATE: 2002-06-20  
; PRIOR APPLICATION NUMBER: U.S. 60/407,006  
; PRIOR FILING DATE: 2002-08-29  
; NUMBER OF SEQ ID NOS: 98  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 31  
; LENGTH: 17  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-600-816-31

Query Match 100.0%; Score 17; DB 8; Length 17;  
Best Local Similarity 100.0%; Pred. No. 75;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGCCGAG 17  
| | | | | | | | | | | | | | | | | |  
Db 1 GCCCAGCTTGCCGAG 17

RESULT 2

US-10-600-816-30  
; Sequence 30, Application US/10600816  
; Publication No. US20040121362A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: IDENTIFICATION AND MODULATION OF A G-PROTEIN COUPLED RECEPTOR  
; TITLE OF INVENTION: (GPCR), RAI3, ASSOCIATED WITH CHRONIC OBSTRUCTIVE PULMONARY  
; TITLE OF INVENTION: DISEASE (COPD) AND NF-KB AND E-SELECTIN REGULATION  
; FILE REFERENCE: D0251 NP  
; CURRENT APPLICATION NUMBER: US/10/600,816  
; CURRENT FILING DATE: 2003-06-20  
; PRIOR APPLICATION NUMBER: U.S. 60/390,850  
; PRIOR FILING DATE: 2002-06-20  
; PRIOR APPLICATION NUMBER: U.S. 60/407,006  
; PRIOR FILING DATE: 2002-08-29  
; NUMBER OF SEQ ID NOS: 98  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 30  
; LENGTH: 42  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-600-816-30

Query Match 100.0%; Score 17; DB 8; Length 42;  
Best Local Similarity 100.0%; Pred. No. 66;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGCCGAG 17  
| | | | | | | | | | | | | | | | | |  
Db 1 GCCCAGCTTGCCGAG 17

RESULT 3  
US-10-066-543-1937  
; Sequence 1937, Application US/10066543  
; Publication No. US20030087818A1  
; GENERAL INFORMATION:  
; APPLICANT: Jiang, Yugu  
; APPLICANT: Pyle, Ruth A.  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Indrias, Carol Yoseph  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Secrist, Heather  
; APPLICANT: Carter, Darrick  
; APPLICANT: Fanger, Gary R.  
; APPLICANT: Smith, Carole L.  
; APPLICANT: Durham, Margarita  
; APPLICANT: Stolk, John A.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER  
; FILE REFERENCE: 210121.563  
; CURRENT APPLICATION NUMBER: US/10/066,543  
; CURRENT FILING DATE: 2002-01-31  
; NUMBER OF SEQ ID NOS: 3417  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1937  
; LENGTH: 497  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-066-543-1937

Query Match 100.0%; Score 17; DB 6; Length 497;  
Best Local Similarity 100.0%; Pred. No. 48;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGCCGAG 17  
| | | | | | | | | | | | | | | | | |  
Db 344 GCCCAGCTTGCCGAG 360

RESULT 4  
US-10-066-543-1811/c  
; Sequence 1811, Application US/10066543

Publication No. US20030087818A1  
GENERAL INFORMATION:  
APPLICANT: Jiang, Yugu  
APPLICANT: Pyle, Ruth A.  
APPLICANT: Xu, Jiangchun  
APPLICANT: Indrias, Carol Yoseph  
APPLICANT: Lodes, Michael J.  
APPLICANT: Secrist, Heather  
APPLICANT: Carter, Darrick  
APPLICANT: Fanger, Gary R.  
APPLICANT: Smith, Carole L.  
APPLICANT: Durham, Margarita  
APPLICANT: Stolk, John A.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
FILE REFERENCE: 210121.563  
CURRENT APPLICATION NUMBER: US/10/066,543  
PRIORITY FILING DATE: 2002-01-31  
NUMBER OF SEQ ID NOS: 3417  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1811  
LENGTH: 552  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-066-543-1811

Query Match 100.0%; Score 17; DB 6; Length 552;  
Best Local Similarity 100.0%; Pred. No. 47;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCCAGCTTGGCCGAG 17  
|||||

Db 154 GCCCAGCTTGGCCGAG 138

## RESULT 5

US-09-969-034-2222/c  
Sequence 2222, Application US/09969034  
Publication No. US20040110668A1  
GENERAL INFORMATION:  
APPLICANT: Burgess, Christopher C.  
APPLICANT: Astle, Jon H.  
APPLICANT: Carroll, Eddie III  
APPLICANT: Catino, Theodore J.  
APPLICANT: Dwivedi, Poornima  
APPLICANT: Molino, Gary A.  
APPLICANT: Thiagalingam, Arunthathi  
APPLICANT: Lewis, Marcia E.  
TITLE OF INVENTION: Nucleic Acid Sequences Differentially  
FILE REFERENCE: 1657/1032  
CURRENT APPLICATION NUMBER: US/09/969,034  
CURRENT FILING DATE: 2001-10-02  
PRIOR APPLICATION NUMBER: 60/237,271  
PRIOR FILING DATE: 2000-02-10  
NUMBER OF SEQ ID NOS: 4494  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 2222  
LENGTH: 620  
TYPE: DNA  
ORGANISM: Homo sapiens  
NAME/KEY: misc feature  
LOCATION: 491, 500, 517, 530, 545, 547, 553, 558, 570, 599, 620  
OTHER INFORMATION: n = A,T,C or G  
US-09-969-034-2222

Query Match 100.0%; Score 17; DB 3; Length 620;  
Best Local Similarity 100.0%; Pred. No. 46;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCCAGCTTGGCCGAG 17  
|||||

Db 154 GCCCAGCTTGGCCGAG 138

## RESULT 6

US-09-969-034-3393/c  
Sequence 3393, Application US/09969034  
Publication No. US20040110668A1  
GENERAL INFORMATION:  
APPLICANT: Burgess, Christopher C.  
APPLICANT: Astle, Jon H.  
APPLICANT: Carroll, Eddie III  
APPLICANT: Catino, Theodore J.  
APPLICANT: Dwivedi, Poornima  
APPLICANT: Molino, Gary A.  
APPLICANT: Thiagalingam, Arunthathi  
APPLICANT: Lewis, Marcia E.  
TITLE OF INVENTION: Nucleic Acid Sequences Differentially  
FILE REFERENCE: 1657/1032  
CURRENT APPLICATION NUMBER: US/09/969,034  
CURRENT FILING DATE: 2001-10-02  
PRIOR APPLICATION NUMBER: 60/237,271  
PRIOR FILING DATE: 2000-02-10  
NUMBER OF SEQ ID NOS: 4494  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 3393  
LENGTH: 634  
TYPE: DNA  
ORGANISM: Homo sapiens  
NAME/KEY: misc feature  
LOCATION: 363, 470, 472, 516, 523, 531, 551, 567, 573, 588, 594, 613,  
LOCATION: 623  
OTHER INFORMATION: n = A,T,C or G  
US-09-969-034-3393

Query Match 100.0%; Score 17; DB 3; Length 634;  
Best Local Similarity 100.0%; Pred. No. 46;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCCAGCTTGGCCGAG 17  
|||||

Db 158 GCCCAGCTTGGCCGAG 142

## RESULT 7

US-10-125-968-701/c  
Sequence 701, Application US/10125968  
Publication No. US20030215805A1  
GENERAL INFORMATION:  
APPLICANT: Lillie, James  
APPLICANT: Palermo, Adam  
APPLICANT: Wang, Youzhen  
APPLICANT: Steinmann, Kathleen  
APPLICANT: Elias, Josh  
APPLICANT: Mertens, Maureen  
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND  
TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
FILE REFERENCE: MRI-032  
CURRENT APPLICATION NUMBER: US/10/125,968  
CURRENT FILING DATE: 2002-04-19  
PRIOR APPLICATION NUMBER: US 60/285,163  
PRIOR FILING DATE: 2001-04-20  
NUMBER OF SEQ ID NOS: 1417  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 701  
LENGTH: 642  
TYPE: DNA  
ORGANISM: Homo sapiens  
NAME/KEY: misc feature  
LOCATION: 72, 76, 77, 99, 101, 423, 445, 468, 489, 557, 566, 633  
US-10-125-968-701/c

Query Match 100.0%; Score 17; DB 3; Length 620;  
Best Local Similarity 100.0%; Pred. No. 46;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCCAGCTTGGCCGAG 17  
|||||

```
; OTHER INFORMATION: n = A,T,C or G
US-10-125-968-701

Query Match      100.0%; Score 17; DB 7; Length 642;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17
Db 184 GCCCAGCGTTGGCCGAG 168

RESULT 8
US-10-712-615-134
; Sequence 134, Application US/10712615
; Publication No. US20040214317A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: A NOVEL HUMAN G-PROTEIN COUPLED RECEPTOR, HGPRBMV8, EXPRESSED
; FILE REFERENCE: D0047A-CIP
; CURRENT FILING DATE: 2003-11-13
; PRIOR APPLICATION NUMBER: US/10/712,615
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: U.S. 09/992,238
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: U.S. 60/248,285
; PRIOR FILING DATE: 2000-11-14
; PRIOR APPLICATION NUMBER: U.S. 60/268,581
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: U.S. 60/308,285
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: U.S. 60/317,166
; PRIOR FILING DATE: 2001-09-04
; NUMBER OF SEQ ID NOS: 134
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 134
; LENGTH: 1071
; TYPE: DNA
; ORGANISM: dna,Homo sapiens
US-10-712-615-134

Query Match      100.0%; Score 17; DB 9; Length 1071;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17
Db 1018 GCCCAGCGTTGGCCGAG 1034

RESULT 9
US-09-866-050A-249
; Sequence 249, Application US/09866050A
; Publication No. US20030040471A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishanand D.
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; FILE REFERENCE: 11000.1011c4U
; CURRENT FILING DATE: 2001-05-24
; CURRENT APPLICATION NUMBER: US/09/866,050A
; NUMBER OF SEQ ID NOS: 725
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 249
; LENGTH: 1212
; TYPE: DNA
; ORGANISM: Human
US-09-866-050A-249
```

```
Query Match      100.0%; Score 17; DB 3; Length 1212;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17
Db 990 GCCCAGCGTTGGCCGAG 1006

RESULT 10
US-10-152-661-249
; Sequence 249, Application US/10152661
; Publication No. US20030022835A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishanand D.
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; FILE REFERENCE: 11000.1011c5
; CURRENT FILING DATE: 2002-05-20
; PRIOR APPLICATION NUMBER: 09/866,050
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/221,232
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: 60/206,650
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: 09/312,283
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: PCT/NZ99/00051
; PRIOR FILING DATE: 1999-04-29
; PRIOR APPLICATION NUMBER: 09/188,930
; PRIOR FILING DATE: 1998-11-09
; PRIOR APPLICATION NUMBER: 09/069,726
; PRIOR FILING DATE: 1998-04-29
; NUMBER OF SEQ ID NOS: 725
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 249
; LENGTH: 1212
; TYPE: DNA
; ORGANISM: Human
US-10-152-661-249

Query Match      100.0%; Score 17; DB 6; Length 1212;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17
Db 990 GCCCAGCGTTGGCCGAG 1006

RESULT 11
US-10-313-542-223
; Sequence 223, Application US/10313542
; Publication No. US20030120057A1
; GENERAL INFORMATION:
; APPLICANT: Roopa, Reddy
; APPLICANT: Guegler, Karl, J.
; APPLICANT: Au-Young, Janice
; TITLE OF INVENTION: COMPOSITION FOR DETECTION OF GENES ENCODING MEMBRANE-ASSOCIATED P
; FILE REFERENCE: PA-0013 US
; CURRENT FILING DATE: 2002-12-05
; CURRENT APPLICATION NUMBER: US/10/313,542
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/118,318
; PRIOR FILING DATE: 1999-02-01
```



NUMBER OF SEQ ID NOS: 305  
SOFTWARE: PERL Program  
SEQ ID NO 223  
LENGTH: 1228  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: Incyte ID No. US20030120057A1 2681738CTI  
US-10-313-542-223

Query Match 100.0%; Score 17; DB 7; Length 1228;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCCGAG 17  
|||||  
Db 541 GCCCAGCTTGGCCGAG 557

## RESULT 12

US-11-060-756-2418  
Sequence 2418, Application US/11060756  
Publication No. US20050221354A1  
GENERAL INFORMATION:  
APPLICANT: Wyeth  
TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug  
FILE REFERENCE: AM101083 (031896-042000)  
CURRENT APPLICATION NUMBER: US/11/060,756  
CURRENT FILING DATE: 2005-02-18  
NUMBER OF SEQ ID NOS: 303284  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 2418  
LENGTH: 1400  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-11-060-756-2418

Query Match 100.0%; Score 17; DB 13; Length 1400;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCCGAG 17  
|||||  
Db 215 GCCCAGCTTGGCCGAG 231

## RESULT 13

US-11-060-756-6690  
Sequence 6690, Application US/11060756  
Publication No. US20050221354A1  
GENERAL INFORMATION:  
APPLICANT: Wyeth  
TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug  
FILE REFERENCE: AM101083 (031896-042000)  
CURRENT APPLICATION NUMBER: US/11/060,756  
CURRENT FILING DATE: 2005-02-18  
NUMBER OF SEQ ID NOS: 303284  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 6690  
LENGTH: 1400  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-11-060-756-6690

Query Match 100.0%; Score 17; DB 13; Length 1400;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCCGAG 17  
|||||  
Db 215 GCCCAGCTTGGCCGAG 231

RESULT 14  
US-10-936-626-64  
Sequence 64, Application US/10936626  
Publication No. US20050106644A1  
GENERAL INFORMATION:  
APPLICANT: Cairns, Belinda  
APPLICANT: Chen, Ruihuan  
APPLICANT: Frantz, Gretchen  
APPLICANT: Hillan, Kenneth J.  
APPLICANT: Koepfen, Hartmut  
APPLICANT: Phillips, Heidi S.  
APPLICANT: Polakis, Paul  
APPLICANT: Spencer, Susan D.  
APPLICANT: Smith, Victoria  
APPLICANT: Williams, P. Mickey  
APPLICANT: Wu, Thomas D.  
APPLICANT: Zhang, Zemin

TITLE OF INVENTION: Compositions and Methods for the Diagnosis and  
TITLE OF INVENTION: Treatment of Tumor  
FILE REFERENCE: P5001R1P1  
CURRENT APPLICATION NUMBER: US/10/936,626  
CURRENT FILING DATE: 2004-09-08  
PRIOR APPLICATION NUMBER: US 10/872,991  
PRIOR FILING DATE: 2004-06-21  
PRIOR APPLICATION NUMBER: US 10/872,972  
PRIOR FILING DATE: 2004-06-21  
PRIOR APPLICATION NUMBER: US 10/241,220  
PRIOR FILING DATE: 2002-09-11  
PRIOR APPLICATION NUMBER: US 10/177,488  
PRIOR FILING DATE: 2002-06-19  
PRIOR APPLICATION NUMBER: US 60/299,500  
PRIOR FILING DATE: 2001-06-20  
PRIOR APPLICATION NUMBER: US 60/301,880  
PRIOR FILING DATE: 2001-06-29  
PRIOR APPLICATION NUMBER: US 60/323,268  
PRIOR FILING DATE: 2001-09-18  
PRIOR APPLICATION NUMBER: US 60/557,116  
PRIOR FILING DATE: 2004-03-26  
PRIOR APPLICATION NUMBER: US 60/598,899  
PRIOR FILING DATE: 2004-08-04  
NUMBER OF SEQ ID NOS: 154  
SEQ ID NO 64  
LENGTH: 1460  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-936-626-64

Query Match 100.0%; Score 17; DB 10; Length 1460;  
Best Local Similarity 100.0%; Pred. No. 41;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCCGAG 17  
|||||  
Db 1145 GCCCAGCTTGGCCGAG 1161

## RESULT 15

US-10-938-061-64  
Sequence 64, Application US/10938061  
Publication No. US20050107595A1  
GENERAL INFORMATION:  
APPLICANT: Cairns, Belinda  
APPLICANT: Chen, Ruihuan  
APPLICANT: Frantz, Gretchen  
APPLICANT: Hillan, Kenneth J.  
APPLICANT: Koepfen, Hartmut  
APPLICANT: Phillips, Heidi S.  
APPLICANT: Polakis, Paul

APPLICANT: Spencer, Susan D.  
APPLICANT: Smith, Victoria  
APPLICANT: Williams, P. Mickey  
APPLICANT: Wu, Thomas D.  
APPLICANT: Zhang, Zemin  
APPLICANT: Sakanaka, Chie  
APPLICANT: Chuntharapai, Anan  
APPLICANT: Reed Chae J.  
TITLE OF INVENTION: Compositions and Methods for the Diagnosis and  
TREATMENT OF Tumor  
FILE REFERENCE: P5001RLP1B  
CURRENT APPLICATION NUMBER: US/10/938,061  
CURRENT FILING DATE: 2004-09-10  
PRIOR APPLICATION NUMBER: US 10/872,991  
PRIOR FILING DATE: 2004-06-21  
PRIOR APPLICATION NUMBER: US 10/872,972  
PRIOR FILING DATE: 2004-06-21  
PRIOR APPLICATION NUMBER: US 10/241,220  
PRIOR FILING DATE: 2002-09-11  
PRIOR APPLICATION NUMBER: US 10/177,488  
PRIOR FILING DATE: 2002-06-19  
PRIOR APPLICATION NUMBER: US 60/299,500  
PRIOR FILING DATE: 2001-06-20  
PRIOR APPLICATION NUMBER: US 60/301,880  
PRIOR FILING DATE: 2001-06-29  
PRIOR APPLICATION NUMBER: US 60/323,268  
PRIOR FILING DATE: 2001-09-18  
PRIOR APPLICATION NUMBER: US 60/557,116  
PRIOR FILING DATE: 2004-03-26  
PRIOR APPLICATION NUMBER: US 60/598,899  
PRIOR FILING DATE: 2004-08-04  
NUMBER OF SEQ ID NOS: 154  
SEQ ID NO 64  
LENGTH: 1460  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-938-061-64

Query Match 100.0%; Score 17; DB 10; Length 1460;  
Best Local Similarity 100.0%; Pred. No. 41;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCCGAG 17  
|||||  
Db 1145 GCCCAGCTTGGCCGAG 1161

RESULT 16  
US-10-224-289-5  
Sequence 5, Application US/10224289  
Publication No. US20030207288A1  
GENERAL INFORMATION:  
APPLICANT: LEWIN, DAVID A.  
APPLICANT: STEWART, TIMOTHY A.  
TITLE OF INVENTION: GPCR-LIKE RETINOIC ACID-INDUCED GENE 1 PROTEIN AND  
METHOD OF INVENTION: NUCLEIC ACID  
FILE REFERENCE: 9800081-0085  
CURRENT APPLICATION NUMBER: US/10/224,289  
CURRENT FILING DATE: 2002-08-20  
PRIOR APPLICATION NUMBER: 60/313,940  
PRIOR FILING DATE: 2001-08-20  
NUMBER OF SEQ ID NOS: 20  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 5  
LENGTH: 1619  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-224-289-5

Query Match 100.0%; Score 17; DB 7; Length 1619;  
Best Local Similarity 100.0%; Pred. No. 41;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCCGAG 17  
|||||  
Db 1137 GCCCAGCTTGGCCGAG 1153

RESULT 17  
US-10-935-190-43  
Sequence 43, Application US/10935190  
Publication No. US20050037466A1  
GENERAL INFORMATION:  
APPLICANT: INCYTE GENOMICS, INC.  
APPLICANT: AU-YOUNG, Janice  
APPLICANT: BANDMAN, Olga  
APPLICANT: TANG, Y. Tom  
APPLICANT: YUE, Henry  
APPLICANT: AZIMZAI, Valda  
APPLICANT: BURFORD, Neil  
APPLICANT: BAUGHN, Mariah R.  
APPLICANT: LU, Dyoung Aina M.  
APPLICANT: HILLMAN, Jennifer L.  
APPLICANT: PATTERSON, Chandra  
APPLICANT: LAL, Preeti  
TITLE OF INVENTION: RECEPTORS AND ASSOCIATED PROTEINS  
FILE REFERENCE: PE-0726 PCT  
CURRENT APPLICATION NUMBER: US/10/935,190  
CURRENT FILING DATE: 2004-09-08  
PRIOR APPLICATION NUMBER: US/10/031,904  
PRIOR FILING DATE: 2002-01-18  
PRIOR APPLICATION NUMBER: 60/145,232; 60/158,578; 60/165,192  
PRIOR FILING DATE: 1999-07-21; 1999-10-07; 1999-11-12  
NUMBER OF SEQ ID NOS: 44  
SOFTWARE: PERL Program  
SEQ ID NO 43  
LENGTH: 1619  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: Incyte ID No: 2681738CB1  
US-10-935-190-43

Query Match 100.0%; Score 17; DB 9; Length 1619;  
Best Local Similarity 100.0%; Pred. No. 41;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCCGAG 17  
|||||  
Db 1137 GCCCAGCTTGGCCGAG 1153

RESULT 18  
US-10-505-486-196  
Sequence 196, Application US/10505486  
Publication No. US20050118639A1  
GENERAL INFORMATION:  
APPLICANT: Takeda Chemical Industries, Ltd.  
TITLE OF INVENTION: Determination of a ligand  
FILE REFERENCE: P03-0006PCT  
CURRENT APPLICATION NUMBER: US/10/505,486  
CURRENT FILING DATE: 2004-08-20  
PRIOR APPLICATION NUMBER: JP 2002-45728  
PRIOR FILING DATE: 2002-02-22  
PRIOR APPLICATION NUMBER: JP 2002-213949  
PRIOR FILING DATE: 2002-07-23  
PRIOR APPLICATION NUMBER: JP 2002-298237  
PRIOR FILING DATE: 2002-10-11  
NUMBER OF SEQ ID NOS: 233  
SEQ ID NO 196  
LENGTH: 1788  
TYPE: DNA  
ORGANISM: Human  
US-10-505-486-196

```
Query Match      100.0%; Score 17; DB 10; Length 1788;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GCCCAGCGTTGGCCGAG 17
      |||
Db      1018 GCCCAGCGTTGGCCGAG 1034

RESULT 19
US-10-775-920-13
; Sequence 13, Application US/10775920
; Publication No. US20040175744A1
; GENERAL INFORMATION:
; APPLICANT: Merigen Ltd
; TITLE OF INVENTION: POLYNUCLEOTIDE SEQUENCES AND CORRESPONDING ENCODED POLYPEPTIDES
; TITLE OF INVENTION: OF PARTICULAR SECRETED AND MEMBRANE-BOUND PROTEINS OVEREXPRESSED
; FILE REFERENCE: Merigen - 00108
; CURRENT APPLICATION NUMBER: US/10/775,920
; PRIOR FILING DATE: 2004-02-10
; PRIOR APPLICATION NUMBER: US 60/447,900
; PRIOR FILING DATE: 2003-02-13
; NUMBER OF SEQ ID NOS: 385
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 13
; LENGTH: 2297
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-775-920-13

Query Match      100.0%; Score 17; DB 9; Length 2297;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GCCCAGCGTTGGCCGAG 17
      |||
Db      1125 GCCCAGCGTTGGCCGAG 1141

RESULT 20
US-10-224-289-3
; Sequence 3, Application US/10224289
; Publication No. US20030207288A1
; GENERAL INFORMATION:
; APPLICANT: LEWIN, DAVID A.
; TITLE OF INVENTION: GPCR-LIKE RETINOIC ACID-INDUCED GENE 1 PROTEIN AND
; TITLE OF INVENTION: NUCLEIC ACID
; FILE REFERENCE: 9800081-0085
; CURRENT APPLICATION NUMBER: US/10/224,289
; CURRENT FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: 60/313,940
; PRIOR FILING DATE: 2001-08-20
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 2302
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-224-289-3

Query Match      100.0%; Score 17; DB 7; Length 2302;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GCCCAGCGTTGGCCGAG 17
      |||
Db      1117 GCCCAGCGTTGGCCGAG 1133

RESULT 21
US-10-240-425-405
```

```
; Sequence 405, Application US/10240425
; Publication No. US20040033502A1
; GENERAL INFORMATION:
; APPLICANT: Williams, Amanda
; APPLICANT: Boland, Joseph F.
; APPLICANT: Lord, Reginald V.
; APPLICANT: Alvarez, Chris
; APPLICANT: Wetzel, Jon C.
; APPLICANT: Scheff, Uwe
; APPLICANT: Vockley, Joseph G.
; TITLE OF INVENTION: Gene Expression Profiles in Esophageal Tissue
; FILE REFERENCE: 44921-5026
; CURRENT APPLICATION NUMBER: US/10/240,425
; CURRENT FILING DATE: 2002-09-30
; PRIOR APPLICATION NUMBER: PCT/US01/09847
; PRIOR FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: US 60/193,446
; PRIOR FILING DATE: 2000-03-31
; NUMBER OF SEQ ID NOS: 1588
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 405
; LENGTH: 2302
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20040033502A1 AF095448
US-10-240-425-405

Query Match      100.0%; Score 17; DB 8; Length 2302;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GCCCAGCGTTGGCCGAG 17
      |||
Db      1117 GCCCAGCGTTGGCCGAG 1133

RESULT 22
US-10-775-920-9
; Sequence 9, Application US/10775920
; Publication No. US20040175744A1
; GENERAL INFORMATION:
; APPLICANT: Merigen Ltd
; TITLE OF INVENTION: POLYNUCLEOTIDE SEQUENCES AND CORRESPONDING ENCODED POLYPEPTIDES
; TITLE OF INVENTION: OF PARTICULAR SECRETED AND MEMBRANE-BOUND PROTEINS OVEREXPRESSED
; FILE REFERENCE: Merigen - 00108
; CURRENT APPLICATION NUMBER: US/10/775,920
; CURRENT FILING DATE: 2004-02-10
; PRIOR APPLICATION NUMBER: US 60/447,900
; PRIOR FILING DATE: 2003-02-13
; NUMBER OF SEQ ID NOS: 385
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9
; LENGTH: 2302
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-775-920-9

Query Match      100.0%; Score 17; DB 9; Length 2302;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GCCCAGCGTTGGCCGAG 17
      |||
Db      1117 GCCCAGCGTTGGCCGAG 1133

RESULT 23
US-10-510-507-2
; Sequence 2, Application US/10510507
; Publication No. US20050282165A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Terrett, Jonathan A
; TITLE OF INVENTION: DIAGNOSIS OF CARCINOMA USING RAIG1 POLYPEPTIDES
; FILE REFERENCE: 2543-1-039PCT/US
; CURRENT APPLICATION NUMBER: US/10/510,507
; CURRENT FILING DATE: 2004-10-07
; PRIOR APPLICATION NUMBER: GB0208331.9
; PRIOR FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: GB0221538.2
; PRIOR FILING DATE: 2002-09-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 2302
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-510-507-2

Query Match      100.0%; Score 17; DB 10; Length 2302;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GCCCAGCGCTTGCCCGAG 17
      |||||||
Db      1117 GCCCAGCGCTTGCCCGAG 1133

RESULT 24
US-10-775-920-12
; Sequence 12, Application US/10775920
; Publication No. US20040175744A1
; GENERAL INFORMATION:
; APPLICANT: MerGen Ltd
; TITLE OF INVENTION: POLYNUCLEOTIDE SEQUENCES AND CORRESPONDING ENCODED POLYPEPTIDES
; TITLE OF INVENTION: OF PARTICULAR SECRETED AND MEMBRANE-BOUND PROTEINS OVEREXPRESSED
; TITLE OF INVENTION: IN CERTAIN CANCERS
; FILE REFERENCE: MerGen - 0010B
; CURRENT APPLICATION NUMBER: US/10/775,920
; CURRENT FILING DATE: 2004-02-10
; PRIOR APPLICATION NUMBER: US 60/447,900
; PRIOR FILING DATE: 2003-02-13
; NUMBER OF SEQ ID NOS: 385
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12
; LENGTH: 2305
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-775-920-12

Query Match      100.0%; Score 17; DB 9; Length 2305;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GCCCAGCGCTTGCCCGAG 17
      |||||||
Db      1121 GCCCAGCGCTTGCCCGAG 1137

RESULT 25
US-10-176-847-59
; Sequence 59, Application US/10176847
; Publication No. US20030068636A1
; GENERAL INFORMATION:
; APPLICANT: Veiby, Petter Ole
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF BREAST
; TITLE OF INVENTION: AND OVARIAN CANCER
; FILE REFERENCE: MRI-039
; CURRENT APPLICATION NUMBER: US/10/176,847
; CURRENT FILING DATE: 2002-06-21
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 59
; LENGTH: 2316
```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-176-847-59

Query Match      100.0%; Score 17; DB 6; Length 2316;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GCCCAGCGCTTGCCCGAG 17
      |||||||
Db      1123 GCCCAGCGCTTGCCCGAG 1139

RESULT 26
US-11-080-991-59
; Sequence 59, Application US/11080991
; Publication No. US20050266437A1
; GENERAL INFORMATION:
; APPLICANT: Veiby, Petter Ole
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF BREAST
; TITLE OF INVENTION: AND OVARIAN CANCER
; FILE REFERENCE: MRI-039
; CURRENT APPLICATION NUMBER: US/11/080,991
; CURRENT FILING DATE: 2005-03-11
; PRIOR APPLICATION NUMBER: US/10/176,847
; PRIOR FILING DATE: 2002-06-21
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 59
; LENGTH: 2316
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-080-991-59

Query Match      100.0%; Score 17; DB 15; Length 2316;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GCCCAGCGCTTGCCCGAG 17
      |||||||
Db      1123 GCCCAGCGCTTGCCCGAG 1139

RESULT 27
US-10-775-920-11
; Sequence 11, Application US/10775920
; Publication No. US20040175744A1
; GENERAL INFORMATION:
; APPLICANT: MerGen Ltd
; TITLE OF INVENTION: POLYNUCLEOTIDE SEQUENCES AND CORRESPONDING ENCODED POLYPEPTIDES
; TITLE OF INVENTION: OF PARTICULAR SECRETED AND MEMBRANE-BOUND PROTEINS OVEREXPRESSED
; TITLE OF INVENTION: IN CERTAIN CANCERS
; FILE REFERENCE: MerGen - 0010B
; CURRENT APPLICATION NUMBER: US/10/775,920
; CURRENT FILING DATE: 2004-02-10
; PRIOR APPLICATION NUMBER: US 60/447,900
; PRIOR FILING DATE: 2003-02-13
; NUMBER OF SEQ ID NOS: 385
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11
; LENGTH: 2446
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-775-920-11

Query Match      100.0%; Score 17; DB 9; Length 2446;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GCCCAGCGCTTGCCCGAG 17
      |||||||
Db      1271 GCCCAGCGCTTGCCCGAG 1287
```

RESULT 28  
US-10-225-567A-453  
; Sequence 453, Application US/10225567A  
; Publication No. US20030113798A1  
; GENERAL INFORMATION:  
; APPLICANT: LifeSpan Biosciences  
; APPLICANT: Brown, Joseph P.  
; APPLICANT: Burner, Glenna C.  
; APPLICANT: Roush, Christine L.  
; TITLE OF INVENTION: ANTIGENIC PEPTIDES AND ANTIBODIES FOR G PROTEIN-COUPLED RECEPTORS  
; FILE REFERENCE: 1920-4-4  
; CURRENT APPLICATION NUMBER: US/10/225,567A  
; CURRENT FILING DATE: 2001-12-19  
; PRIOR APPLICATION NUMBER: 60/257,144  
; PRIOR FILING DATE: 2000-12-19  
; NUMBER OF SEQ ID NOS: 2292  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 453  
; LENGTH: 2456  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-225-567A-453

Query Match 100.0%; Score 17; DB 6; Length 2456;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCCAGCTTGCCGAG 17  
|||||  
DB 1271 GCCCAGCTTGCCGAG 1287

RESULT 29  
US-10-269-909-63  
; Sequence 63, Application US/10269909  
; Publication No. US20030180747A1  
; GENERAL INFORMATION:  
; APPLICANT: HRUBAN, RALPH H.  
; APPLICANT: ARGANI, PEDRAM  
; APPLICANT: IACOBUZIO-DONAHUE, CHRISTINE  
; APPLICANT: MAITRA, ANIRBAN  
; TITLE OF INVENTION: PANCREATIC CANCER DIAGNOSIS AND THERAPIES  
; FILE REFERENCE: 58303(71699)  
; CURRENT APPLICATION NUMBER: US/10/269,909  
; CURRENT FILING DATE: 2003-10-11  
; PRIOR APPLICATION NUMBER: 60/328,609  
; PRIOR FILING DATE: 2001-10-11  
; PRIOR APPLICATION NUMBER: 60/332,754  
; PRIOR FILING DATE: 2001-11-19  
; NUMBER OF SEQ ID NOS: 87  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 63  
; LENGTH: 2456  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-269-909-63

Query Match 100.0%; Score 17; DB 7; Length 2456;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCCAGCTTGCCGAG 17  
|||||  
DB 1271 GCCCAGCTTGCCGAG 1287

RESULT 30  
US-10-269-909-64  
; Sequence 64, Application US/10269909  
; Publication No. US20030180747A1  
; GENERAL INFORMATION:

; APPLICANT: HRUBAN, RALPH H.  
; APPLICANT: ARGANI, PEDRAM  
; APPLICANT: IACOBUZIO-DONAHUE, CHRISTINE  
; APPLICANT: MAITRA, ANIRBAN  
; TITLE OF INVENTION: PANCREATIC CANCER DIAGNOSIS AND THERAPIES  
; FILE REFERENCE: 58303(71699)  
; CURRENT APPLICATION NUMBER: US/10/269,909  
; CURRENT FILING DATE: 2003-10-11  
; PRIOR APPLICATION NUMBER: 60/328,609  
; PRIOR FILING DATE: 2001-10-11  
; PRIOR APPLICATION NUMBER: 60/332,754  
; PRIOR FILING DATE: 2001-11-19  
; NUMBER OF SEQ ID NOS: 87  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 64  
; LENGTH: 2456  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-269-909-64

Query Match 100.0%; Score 17; DB 7; Length 2456;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCCAGCTTGCCGAG 17  
|||||  
DB 1271 GCCCAGCTTGCCGAG 1287

RESULT 31  
US-10-295-027-619  
; Sequence 619, Application US/10295027  
; Publication No. US20030232350A1  
; GENERAL INFORMATION:  
; APPLICANT: Afar, Daniel  
; APPLICANT: Aziz, Natasha  
; APPLICANT: Ginsberg, Wendy M.  
; APPLICANT: Gish, Kurt C.  
; APPLICANT: Glynn, Richard  
; APPLICANT: Hevezi, Peter A.  
; APPLICANT: Mack, David H.  
; APPLICANT: Murray, Richard  
; APPLICANT: Watson, Susan R.  
; APPLICANT: Eos Biotechnology, Inc.  
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and  
; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer  
; FILE REFERENCE: 018501-012500US  
; CURRENT APPLICATION NUMBER: US/10/295,027  
; CURRENT FILING DATE: 2002-11-13  
; PRIOR APPLICATION NUMBER: US 09/663,733  
; PRIOR FILING DATE: 2000-09-15  
; PRIOR APPLICATION NUMBER: US 60/350,666  
; PRIOR FILING DATE: 2001-11-13  
; PRIOR APPLICATION NUMBER: US 60/335,394  
; PRIOR FILING DATE: 2001-11-15  
; PRIOR APPLICATION NUMBER: US 60/332,464  
; PRIOR FILING DATE: 2001-11-21  
; PRIOR APPLICATION NUMBER: US 60/334,393  
; PRIOR FILING DATE: 2001-11-29  
; PRIOR APPLICATION NUMBER: US 60/340,376  
; PRIOR FILING DATE: 2001-12-14  
; PRIOR APPLICATION NUMBER: US 60/347,211  
; PRIOR FILING DATE: 2002-01-08  
; PRIOR APPLICATION NUMBER: US 60/347,349  
; PRIOR FILING DATE: 2002-01-10  
; PRIOR APPLICATION NUMBER: US 60/355,250  
; PRIOR FILING DATE: 2002-02-08  
; PRIOR APPLICATION NUMBER: US 60/356,714  
; PRIOR FILING DATE: 2002-02-13  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 1386  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 619

```
; LENGTH: 2456
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-295-027-619

Query Match      100.0%; Score 17; DB 7; Length 2456;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGCTTGGCCGAG 17
Db 1271 GCCCAGCGCTTGGCCGAG 1287

RESULT 32
US-10-600-816-2
; Sequence 2, Application US/10600816
; Publication No. US20040121362A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: IDENTIFICATION AND MODULATION OF A G-PROTEIN COUPLED RECEPTOR
; TITLE OF INVENTION: (GPCR), RA13, ASSOCIATED WITH CHRONIC OBSTRUCTIVE PULMONARY
; TITLE OF INVENTION: DISEASE (COPD) AND NF-KB AND E-SELECTIN REGULATION
; FILE REFERENCE: D0251 NP
; CURRENT APPLICATION NUMBER: US/10/600,816
; CURRENT FILING DATE: 2003-06-20
; PRIOR APPLICATION NUMBER: U.S. 60/390,850
; PRIOR FILING DATE: 2002-06-20
; PRIOR APPLICATION NUMBER: U.S. 60/407,006
; PRIOR FILING DATE: 2002-08-29
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 2456
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-600-816-2

Query Match      100.0%; Score 17; DB 8; Length 2456;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGCTTGGCCGAG 17
Db 1271 GCCCAGCGCTTGGCCGAG 1287

RESULT 33
US-10-600-816-18
; Sequence 18, Application US/10600816
; Publication No. US20040121362A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: IDENTIFICATION AND MODULATION OF A G-PROTEIN COUPLED RECEPTOR
; TITLE OF INVENTION: (GPCR), RA13, ASSOCIATED WITH CHRONIC OBSTRUCTIVE PULMONARY
; TITLE OF INVENTION: DISEASE (COPD) AND NF-KB AND E-SELECTIN REGULATION
; FILE REFERENCE: D0251 NP
; CURRENT APPLICATION NUMBER: US/10/600,816
; CURRENT FILING DATE: 2003-06-20
; PRIOR APPLICATION NUMBER: U.S. 60/390,850
; PRIOR FILING DATE: 2002-06-20
; PRIOR APPLICATION NUMBER: U.S. 60/407,006
; PRIOR FILING DATE: 2002-08-29
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 18
; LENGTH: 2456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: RA13 Polymorphic Allele Summary Sequence.
; FEATURE:
; NAME/KEY: misc_feature
```

```
; LOCATION: (112)..(112)
; OTHER INFORMATION: wherein "n" equals either G or A.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (364)..(364)
; OTHER INFORMATION: wherein "n" equals either C or T.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (511)..(511)
; OTHER INFORMATION: wherein "n" equals either C or T.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (523)..(523)
; OTHER INFORMATION: wherein "n" equals either C or T.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (605)..(605)
; OTHER INFORMATION: wherein "n" equals either A or G.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (797)..(797)
; OTHER INFORMATION: wherein "n" equals either A or G.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1111)..(1111)
; OTHER INFORMATION: wherein "n" equals either T or C.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1173)..(1173)
; OTHER INFORMATION: wherein "n" equals either A or G.
US-10-600-816-18

Query Match      100.0%; Score 17; DB 8; Length 2456;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGCTTGGCCGAG 17
Db 1271 GCCCAGCGCTTGGCCGAG 1287

RESULT 34
US-10-775-920-10
; Sequence 10, Application US/10775920
; Publication No. US2004017544A1
; GENERAL INFORMATION:
; APPLICANT: Merigen Ltd
; TITLE OF INVENTION: POLYNUCLEOTIDE SEQUENCES AND CORRESPONDING ENCODED POLYPEPTIDES
; TITLE OF INVENTION: OF PARTICULAR SECRETED AND MEMBRANE-BOUND PROTEINS OVEREXPRESSED
; FILE REFERENCE: Merigen - 0010B
; CURRENT APPLICATION NUMBER: US/10/775,920
; CURRENT FILING DATE: 2004-02-10
; PRIOR APPLICATION NUMBER: US 60/447,900
; PRIOR FILING DATE: 2003-02-13
; NUMBER OF SEQ ID NOS: 385
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 10
; LENGTH: 2456
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-775-920-10

Query Match      100.0%; Score 17; DB 9; Length 2456;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCGCTTGGCCGAG 17
Db 1271 GCCCAGCGCTTGGCCGAG 1287

RESULT 35
```

US-10-936-626-40  
; Sequence 40, Application US/10936626  
; Publication No. US20050106644A1  
; GENERAL INFORMATION:  
; APPLICANT: Cairns, Belinda  
; APPLICANT: Chen, Ruihuan  
; APPLICANT: Frantz, Gretchen  
; APPLICANT: Hillan, Kenneth J.  
; APPLICANT: Koepfen, Hartmut  
; APPLICANT: Phillips, Heidi S.  
; APPLICANT: Polakis, Paul  
; APPLICANT: Spencer, Susan D.  
; APPLICANT: Smith, Victoria  
; APPLICANT: Williams, P. Mickey  
; APPLICANT: Wu, Thomas D.  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: Compositions and Methods for the Diagnosis and  
; TITLE OF INVENTION: Treatment of Tumor  
; FILE REFERENCE: P5001R1P1  
; CURRENT APPLICATION NUMBER: US/10/936,626  
; PRIOR FILING DATE: 2004-09-08  
; PRIOR APPLICATION NUMBER: US 10/872,991  
; PRIOR FILING DATE: 2004-06-21  
; PRIOR APPLICATION NUMBER: US 10/872,972  
; PRIOR FILING DATE: 2004-06-21  
; PRIOR APPLICATION NUMBER: US 10/241,220  
; PRIOR FILING DATE: 2002-09-11  
; PRIOR APPLICATION NUMBER: US 10/177,488  
; PRIOR FILING DATE: 2002-06-19  
; PRIOR APPLICATION NUMBER: US 60/299,500  
; PRIOR FILING DATE: 2001-06-20  
; PRIOR APPLICATION NUMBER: US 60/301,880  
; PRIOR FILING DATE: 2001-06-29  
; PRIOR APPLICATION NUMBER: US 60/323,268  
; PRIOR FILING DATE: 2001-09-18  
; PRIOR APPLICATION NUMBER: US 60/557,116  
; PRIOR FILING DATE: 2004-03-26  
; PRIOR APPLICATION NUMBER: US 60/598,899  
; PRIOR FILING DATE: 2004-08-04  
; NUMBER OF SEQ ID NOS: 154  
; SEQ ID NO 40  
; LENGTH: 2456  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-936-626-40

Query Match 100.0%; Score 17; DB 10; Length 2456;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GCCCAGCGTTGCCGAG 17  
|||||  
Db 1271 GCCCAGCGTTGCCGAG 1287

RESULT 36  
US-10-938-061-40  
; Sequence 40, Application US/10938061  
; Publication No. US20050107595A1  
; GENERAL INFORMATION:  
; APPLICANT: Cairns, Belinda  
; APPLICANT: Chen, Ruihuan  
; APPLICANT: Frantz, Gretchen  
; APPLICANT: Hillan, Kenneth J.  
; APPLICANT: Koepfen, Hartmut  
; APPLICANT: Phillips, Heidi S.  
; APPLICANT: Polakis, Paul  
; APPLICANT: Spencer, Susan D.  
; APPLICANT: Smith, Victoria  
; APPLICANT: Williams, P. Mickey  
; APPLICANT: Wu, Thomas D.  
; APPLICANT: Zhang, Zemin  
; APPLICANT: Sakanaka, Chie

; APPLICANT: Chuntharapai, Anan  
; APPLICANT: Reed Chae J.  
; TITLE OF INVENTION: Compositions and Methods for the Diagnosis and  
; TITLE OF INVENTION: Treatment of Tumor  
; FILE REFERENCE: P5001R1P1B  
; CURRENT APPLICATION NUMBER: US/10/938,061  
; PRIOR FILING DATE: 2004-09-10  
; PRIOR APPLICATION NUMBER: US 10/872,991  
; PRIOR FILING DATE: 2004-06-21  
; PRIOR APPLICATION NUMBER: US 10/872,972  
; PRIOR FILING DATE: 2004-06-21  
; PRIOR APPLICATION NUMBER: US 10/241,220  
; PRIOR FILING DATE: 2002-09-11  
; PRIOR APPLICATION NUMBER: US 10/177,488  
; PRIOR FILING DATE: 2002-06-19  
; PRIOR APPLICATION NUMBER: US 60/299,500  
; PRIOR FILING DATE: 2001-06-20  
; PRIOR APPLICATION NUMBER: US 60/301,880  
; PRIOR FILING DATE: 2001-06-29  
; PRIOR APPLICATION NUMBER: US 60/323,268  
; PRIOR FILING DATE: 2001-09-18  
; PRIOR APPLICATION NUMBER: US 60/557,116  
; PRIOR FILING DATE: 2004-03-26  
; PRIOR APPLICATION NUMBER: US 60/598,899  
; PRIOR FILING DATE: 2004-08-04  
; NUMBER OF SEQ ID NOS: 154  
; SEQ ID NO 40  
; LENGTH: 2456  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-938-061-40

Query Match 100.0%; Score 17; DB 10; Length 2456;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GCCCAGCGTTGCCGAG 17  
|||||  
Db 1271 GCCCAGCGTTGCCGAG 1287

RESULT 37  
US-11-169-041-32  
; Sequence 32, Application US/11169041  
; Publication No. US20060019284A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: IDENTIFICATION OF POLYNUCLEOTIDES FOR PREDICTING ACTIVITY OF  
; TITLE OF INVENTION: COMPOUNDS THAT INTERACT WITH AND/OR MODULATE PROTEIN TYROSINE  
; TITLE OF INVENTION: KINASES AND/OR PROTEIN TYROSINE KINASE PATHWAYS IN LUNG CANCER  
; TITLE OF INVENTION: CELLS  
; FILE REFERENCE: 10001 NP  
; CURRENT APPLICATION NUMBER: US/11/169,041  
; CURRENT FILING DATE: 2005-06-28  
; PRIOR APPLICATION NUMBER: 60/584,405  
; PRIOR FILING DATE: 2004-06-30  
; NUMBER OF SEQ ID NOS: 527  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 32  
; LENGTH: 2456  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-169-041-32

Query Match 100.0%; Score 17; DB 16; Length 2456;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GCCCAGCGTTGCCGAG 17  
|||||  
Db 1271 GCCCAGCGTTGCCGAG 1287

```

RESULT 38
US-10-264-049-834
; Sequence 834, Application US/10264049
; Publication No. US20040005579A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA133PI
; CURRENT APPLICATION NUMBER: US/10/264,049
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/18569
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 4360
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 834
; LENGTH: 2593
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-264-049-834

Query Match 100.0%; Score 17; DB 7; Length 2593;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGCCGAG 17
Db 1383 GCCCAGCTTGCCGAG 1399

RESULT 39
US-10-198-846-10424
; Sequence 10424, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10424
; LENGTH: 4239
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1, 4238, 4239
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-10424

Query Match 100.0%; Score 17; DB 6; Length 4239;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGCCGAG 17
Db 1679 GCCCAGCTTGCCGAG 1695

RESULT 40
US-10-335-977-1640/c
; Sequence 1640, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA

```

```

; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 742-4214
; INFORMATION FOR SEQ ID NO: 1640:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1020 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...1020
; SEQUENCE DESCRIPTION: SEQ ID NO: 1640:
US-10-335-977-1640

Query Match 90.6%; Score 15.4; DB 8; Length 1020;
Best Local Similarity 94.1%; Pred. No. 3e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGCCGAG 17
Db 981 GCCCAGCTTGCCGAG 965

RESULT 41
US-10-335-977-1641/c
; Sequence 1641, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA

```



ZIP: 02109-1875  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: Windows NT 4.0  
SOFTWARE: UNIX  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/335,977  
FILING DATE: 30-Dec-2002  
APPLICATION DATA:  
APPLICATION NUMBER: 08/993,002  
FILING DATE: 17-Dec-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Mandragouras, Amy E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: GTN-018  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)742-4214  
INFORMATION FOR SEQ ID NO: 1641:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1050 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Helicobacter pylori  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (8) LOCATION 1...1050  
SEQUENCE DESCRIPTION: SEQ ID NO: 1641:

US-10-335-977-1641

Query Match 90.6%; Score 15.4; DB 8; Length 1050;  
Best Local Similarity 94.1%; Pred. No. 3e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCCAGCGTTGGCCGAG 17  
|||||  
DB 1014 GCCCAGCGTTGGCCGAG 998

## RESULT 42

US-10-956-157-41  
; Sequence 41, Application US/10956157  
; Publication No. US20050118625A1  
; GENERAL INFORMATION:

; APPLICANT: Wyeth  
; APPLICANT: Mounts, William  
; TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH  
; FILE REFERENCE: 031896-043000 (AM 101081)  
; CURRENT APPLICATION NUMBER: US/10/956,157  
; CURRENT FILING DATE: 2004-10-04  
; NUMBER OF SEQ ID NOS: 319805  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 41  
; LENGTH: 1173  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-956-157-41

Query Match 90.6%; Score 15.4; DB 10; Length 1173;  
Best Local Similarity 94.1%; Pred. No. 3e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCCAGCGTTGGCCGAG 17  
|||||  
DB 228 GCCCAGCGTTGGCCGAG 244

## RESULT 43

US-10-956-157-5276  
; Sequence 5276, Application US/10956157  
; Publication No. US20050118625A1  
; GENERAL INFORMATION:

; APPLICANT: Wyeth  
; APPLICANT: Mounts, William  
; TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH  
; FILE REFERENCE: 031896-043000 (AM 101081)  
; CURRENT APPLICATION NUMBER: US/10/956,157  
; CURRENT FILING DATE: 2004-10-04  
; NUMBER OF SEQ ID NOS: 319805  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 5276  
; LENGTH: 1173  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-956-157-5276

Query Match 90.6%; Score 15.4; DB 10; Length 1173;  
Best Local Similarity 94.1%; Pred. No. 3e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCCAGCGTTGGCCGAG 17  
|||||  
DB 228 GCCCAGCGTTGGCCGAG 244

## RESULT 44

US-10-398-221-3404/c  
; Sequence 3404, Application US/10398221  
; Publication No. US20040018514A1  
; GENERAL INFORMATION:

; APPLICANT: KUNST, Frederik  
; APPLICANT: GLASER, Philippe  
; TITLE OF INVENTION: Listeria innocua, genome and applications  
; FILE REFERENCE: 344 702 - US  
; CURRENT APPLICATION NUMBER: US/10/398,221  
; CURRENT FILING DATE: 2003-03-27  
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061  
; PRIOR FILING DATE: 2001-10-04  
; PRIOR APPLICATION NUMBER: FR 00/12 697  
; PRIOR FILING DATE: 2000-10-04  
; NUMBER OF SEQ ID NOS: 4025  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 3404  
; LENGTH: 2306  
; TYPE: DNA  
; ORGANISM: Listeria monocytogenes 4b  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(end)  
; OTHER INFORMATION: n can be any nucleotide: a,g,c or t/u  
US-10-398-221-3404

Query Match 90.6%; Score 15.4; DB 8; Length 2306;  
Best Local Similarity 94.1%; Pred. No. 2.7e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCCAGCGTTGGCCGAG 17  
|||||  
DB 834 GCCCAGCGTTGGCCGAG 818

## RESULT 45

US-09-757-716-2  
; Sequence 2, Application US/09757716  
; Patent No. US20010012515A1  
; GENERAL INFORMATION:

; APPLICANT: Magna, Holly  
; APPLICANT: Schaffer, Paul

```
; APPLICANT: Lawton, Michael
; APPLICANT: Yocum, Sue
; APPLICANT: Mitchell, Peter
; APPLICANT: Hutchinson, Nancy
; APPLICANT: Murrey, Lynn E.
; TITLE OF INVENTION: HUMAN NUCLEOTIDE PYROPHOSPHOHYDROLASE-2
; FILE REFERENCE: PF-0420 US
; CURRENT APPLICATION NUMBER: US/09/757,716
; CURRENT FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: US/08/996,083
; PRIOR FILING DATE: 1997-12-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 4183
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(4183)
; OTHER INFORMATION: n = A,T,C or G
; PUBLICATION INFORMATION:
US-09-757-716-2

Query Match          90.6%; Score 15.4; DB 3; Length 4183;
Best Local Similarity 94.1%; Pred. No. 2.5e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17
Db 3254 GCCCAGCGTTGGCCAG 3270

RESULT 46
US-10-723-860-7430
; Sequence 7430, Application US/107233860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziiz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods for Diagnosis of Soft Tissue Sarcoma, Compositions &
; FILE REFERENCE: 03682.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 7430
; LENGTH: 4222
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2438)..(2503)
; OTHER INFORMATION: n is a, c, g, or t
US-10-723-860-7430

Query Match          90.6%; Score 15.4; DB 9; Length 4222;
Best Local Similarity 94.1%; Pred. No. 2.5e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17
Db 3295 GCCCAGCGTTGGCCAG 3311

RESULT 47
US-10-322-281-557
; Sequence 557, Application US/10322281
; Publication No. US20040126762A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc S. Malandro
; TITLE OF INVENTION: Novel Compositions and Methods in Cancer
; FILE REFERENCE: 529452001000
; CURRENT APPLICATION NUMBER: US/10/322,281
; CURRENT FILING DATE: 2002-12-17
; NUMBER OF SEQ ID NOS: 866
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 557
; LENGTH: 63761
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(63761)
; OTHER INFORMATION: n = A,T,C or G
US-10-322-281-557

Query Match          90.6%; Score 15.4; DB 8; Length 63761;
Best Local Similarity 94.1%; Pred. No. 1.7e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17
Db 34522 GCCCAGCGTTGGCCGAG 34538

RESULT 48
US-10-322-281-560
; Sequence 560, Application US/10322281
; Publication No. US20040126762A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc S. Malandro
; TITLE OF INVENTION: Novel Compositions and Methods in Cancer
; FILE REFERENCE: 529452001000
; CURRENT APPLICATION NUMBER: US/10/322,281
; CURRENT FILING DATE: 2002-12-17
; NUMBER OF SEQ ID NOS: 866
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 560
; LENGTH: 68732
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(68732)
; OTHER INFORMATION: n = A,T,C or G
US-10-322-281-560

Query Match          90.6%; Score 15.4; DB 8; Length 68732;
Best Local Similarity 94.1%; Pred. No. 1.7e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17
Db 38569 GCCCAGCGTTGGCCGAG 38585

RESULT 49
US-10-398-221-8
; Sequence 8, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398,221
; CURRENT FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
```

```
; APPLICANT: David W. Morris
; APPLICANT: Marc S. Malandro
; TITLE OF INVENTION: Novel Compositions and Methods in Cancer
; FILE REFERENCE: 529452001000
; CURRENT APPLICATION NUMBER: US/10/322,281
; CURRENT FILING DATE: 2002-12-17
; NUMBER OF SEQ ID NOS: 866
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 557
; LENGTH: 63761
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(63761)
; OTHER INFORMATION: n = A,T,C or G
US-10-322-281-557

Query Match          90.6%; Score 15.4; DB 8; Length 63761;
Best Local Similarity 94.1%; Pred. No. 1.7e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17
Db 34522 GCCCAGCGTTGGCCGAG 34538

RESULT 48
US-10-322-281-560
; Sequence 560, Application US/10322281
; Publication No. US20040126762A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc S. Malandro
; TITLE OF INVENTION: Novel Compositions and Methods in Cancer
; FILE REFERENCE: 529452001000
; CURRENT APPLICATION NUMBER: US/10/322,281
; CURRENT FILING DATE: 2002-12-17
; NUMBER OF SEQ ID NOS: 866
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 560
; LENGTH: 68732
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(68732)
; OTHER INFORMATION: n = A,T,C or G
US-10-322-281-560

Query Match          90.6%; Score 15.4; DB 8; Length 68732;
Best Local Similarity 94.1%; Pred. No. 1.7e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGAG 17
Db 38569 GCCCAGCGTTGGCCGAG 38585

RESULT 49
US-10-398-221-8
; Sequence 8, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398,221
; CURRENT FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
```

;; PRIOR FILING DATE: 2000-10-04  
;; NUMBER OF SEQ ID NOS: 4025  
;; SOFTWARE: PatentIn version 3.0  
;; SEQ ID NO 8  
;; LENGTH: 495269  
;; TYPE: DNA  
;; ORGANISM: Listeria innocua  
;; FEATURE:  
;; NAME/KEY: misc feature  
;; LOCATION: (1)..(end)  
;; OTHER INFORMATION: n can be any nucleotide: a, g, c or t/u  
US-10-398-221-8

Query Match 90.6%; Score 15.4; DB 8; Length 495269;  
Best Local Similarity 94.1%; Pred. No. 1.3e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGCCGAG 17  
Db 400813 GCCCAGCTTGCTGAG 400829

## RESULT 50

US-10-470-565-1/c  
;; Sequence 1, Application US/10470565  
;; Publication No. US20040126870A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Societe des Produits Nestle S.A.  
;; TITLE OF INVENTION: NC2705 - the genome of a Bifidobacterium  
;; FILE REFERENCE: 80290/WO  
;; CURRENT APPLICATION NUMBER: US/10/470,565  
;; CURRENT FILING DATE: 2003-07-29  
;; PRIOR APPLICATION NUMBER: EP 01102050.0  
;; PRIOR FILING DATE: 2001-01-30  
;; NUMBER OF SEQ ID NOS: 2  
;; SOFTWARE: PatentIn version 3.1  
;; SEQ ID NO 1  
;; LENGTH: 2256646  
;; TYPE: DNA  
;; ORGANISM: Bifidobacterium longum  
US-10-470-565-1

Query Match 90.6%; Score 15.4; DB 8; Length 2256646;  
Best Local Similarity 94.1%; Pred. No. 1.1e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGCCGAG 17  
Db 1222897 GCCCAGCTTGCCGAG 1222881

## RESULT 51

US-11-045-004-1/c  
;; Sequence 1, Application US/11045004  
;; Publication No. US20060078901A1  
;; GENERAL INFORMATION:  
;; APPLICANT: BUCHRIESER, CARMEN  
;; APPLICANT: FRANGEUL, LIONEL  
;; APPLICANT: COUVE, ELISABETH  
;; APPLICANT: RUSNIOK, CHRISTOPHE  
;; APPLICANT: FSIHI, HAFIDA  
;; APPLICANT: DEHOUX, PIERRE  
;; APPLICANT: DUSSURGET, OLIVIER  
;; APPLICANT: CHETOUANI, FARID  
;; APPLICANT: NEJARI, HAFED  
;; APPLICANT: GLASER, PHILIPPE  
;; APPLICANT: KUNST, FRANCK  
;; APPLICANT: COSSART, PASCALE  
;; APPLICANT: DANIELS, JUSTIN  
;; APPLICANT: GOEBEL, WERNER  
;; APPLICANT: KREFT, JURGEN  
;; APPLICANT: KUHN, MICHAEL  
;; APPLICANT: NG, EVA

;; APPLICANT: VAZQUEZ-BOLAND, ANTONIO  
;; APPLICANT: DOMINGUEZ-BERNAL, GUSTAVO  
;; APPLICANT: GARRIDO-GARCIA, PATRICIA  
;; APPLICANT: TIERREZ-MARTINEZ, ALBERTO  
;; APPLICANT: AMEND, ALEXANDRA  
;; APPLICANT: CHAKRABORTY, TRINAD  
;; APPLICANT: DOMANN, EUGEN  
;; APPLICANT: HAIN, THORSTEN  
;; APPLICANT: BERCHE, PATRICK  
;; APPLICANT: CHARBIT, ALAIN  
;; APPLICANT: DURANT, LIONEL  
;; APPLICANT: PEREZ-DIAZ, JOSE-CLAUDIO  
;; APPLICANT: BAQUERO, FERNANDO  
;; APPLICANT: GARCIA DEL PORTILLO, FRANCISCO  
;; APPLICANT: GOMEZ-LOPEZ, NURIA  
;; APPLICANT: MADUENIO, ENCARNNA  
;; APPLICANT: PABLOS, BETRIZ DE  
;; APPLICANT: WEHLAND, JURGEN  
;; APPLICANT: KARST, UWE  
;; APPLICANT: ENTIAN, KARL-DIETER  
;; APPLICANT: HAUF, JORG  
;; APPLICANT: ROSE, MATTHIAS  
;; APPLICANT: VOSS, HAMUT  
;; TITLE OF INVENTION: LISTERIA MONOCYTOGENES GENOME, POLYPEPTIDES AND USES  
;; FILE REFERENCE: 05394.0018-02  
;; CURRENT APPLICATION NUMBER: US/11/045,004  
;; CURRENT FILING DATE: 2005-01-28  
;; PRIOR APPLICATION NUMBER: 10/637,657  
;; PRIOR FILING DATE: 2003-08-11  
;; PRIOR APPLICATION NUMBER: 10/257,023  
;; PRIOR FILING DATE: 2002-10-08  
;; PRIOR APPLICATION NUMBER: PCT/FR01/01118  
;; PRIOR FILING DATE: 2001-04-11  
;; PRIOR APPLICATION NUMBER: FR 00/04,629  
;; PRIOR FILING DATE: 2000-04-11  
;; NUMBER OF SEQ ID NOS: 2854  
;; SOFTWARE: PatentIn version 3.3  
;; SEQ ID NO 1  
;; LENGTH: 2944528  
;; TYPE: DNA  
;; ORGANISM: Listeria monocytogenes  
US-11-045-004-1

Query Match 90.6%; Score 15.4; DB 16; Length 2944528;  
Best Local Similarity 94.1%; Pred. No. 1e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGCCGAG 17  
Db 1351366 GCCCAGCTTGCTGAG 1351350

## RESULT 52

US-10-398-221-2058/c  
;; Sequence 2058, Application US/10398221  
;; Publication No. US20040018514A1  
;; GENERAL INFORMATION:  
;; APPLICANT: KUNST, Frederik  
;; APPLICANT: GLASER, Philippe  
;; TITLE OF INVENTION: Listeria innocua, genome and applications  
;; FILE REFERENCE: 344 702 - US  
;; CURRENT APPLICATION NUMBER: US/10/398,221  
;; CURRENT FILING DATE: 2003-03-27  
;; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061  
;; PRIOR FILING DATE: 2001-10-04  
;; PRIOR APPLICATION NUMBER: FR 00/12 697  
;; PRIOR FILING DATE: 2000-10-04  
;; NUMBER OF SEQ ID NOS: 4025  
;; SOFTWARE: PatentIn version 3.0  
;; SEQ ID NO 2058  
;; LENGTH: 3011208  
;; TYPE: DNA  
;; ORGANISM: Listeria innocua

US-10-398-221-2058

```
Query Match      90.6%; Score 15.4; DB 8; Length 3011208;
Best Local Similarity 94.1%; Pred. No. 1e+02;
Matches 16: Conservative 0; Mismatches 1; Indels 0;
```

Qy 1 GCCCAGCTTGGCCGAG 17  
1358141 GCCCAGCTTGGCTGAG 1358125  
Db

RESULT 53

```

US-10-357-930-57611/c
; Sequence 57611, Application US/10357930
; Publication No. US20040259086A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Endege, Wilson
; APPLICANT: Monahan, John
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF
; TITLE OF INVENTION: HUMAN PROSTATE CANCER
; FILE REFERENCE: MRI-007BCN
; CURRENT APPLICATION NUMBER: US/10/357,930
; CURRENT FILING DATE: 2003-02-04
; PRIOR APPLICATION NUMBER: 09/785,276
; PRIOR FILING DATE: 2003-02-16
; PRIOR APPLICATION NUMBER: 60/183,319
; PRIOR FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 60/189,862
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: 60/207,454
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/211,314
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: 60/219,007
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/255,281
; PRIOR FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 62232
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 57611
; LENGTH: 337
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-357-930-57611

```

```
Query Match      88.2%; Score 15; DB 9; Length 337;
Best Local Similarity 100.0%; Pred. No. 5.7e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy 3 CCACGCTTGGCCGAG 17  
Db 216 CCACGCTTGGCCGAG 202

## RESULT 54

```

US-10-425-115-123425/c
; Sequence 123425, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21 (53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 123425
; LENGTH: 404

```

```

; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_44044C.1
US-10-425-115-123425

```

Query Match 88.2%; Score 15; DB 9; Length 404;  
Best Local Similarity 100.0%; Pred. No. 5.6e+02;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCACGCTTGGCCG 15  
pb 136 GCCACGCTTGGCCG 122

RESULT 55

```

US-10-739-930-4894/c
; Sequence 4894, Application US/10739930
; Publication No. US20040216190A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES
; TITLE OF INVENTION: PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53377)B
; CURRENT APPLICATION NUMBER: US/10/739,930
; CURRENT FILING DATE: 2003-12-18
; NUMBER OF SEQ ID NOS: 11088
; SEQ ID NO 4894
; LENGTH: 3303
; TYPE: DNA
; ORGANISM: Triticum aestivum
; FEATURE:
; OTHER INFORMATION: Clone ID: TRIAE-23APR03-CLUSTER413830_1
US-10-739-930-4894

```

```
Query Match      88.2%; Score 15; DB 9; Length 3303;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 15: Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy 2 CCCACGCTTGGCCGA 16  
p6 619 CCCACGCTTGGCCGA 605

RESULT 56

```

US-10-956-157-193814
; Sequence 193814, Application US/10956157
; Publication No. US20050118625A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William
; TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION
; FILE REFERENCE: 031896-043000 (AM 101081)
; CURRENT APPLICATION NUMBER: US/10/956.157
; CURRENT FILING DATE: 2004-10-04
; NUMBER OF SEQ ID NOS: 319805
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 193814
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Probe Sequence
US-10-956-157-193814

```

Query Match	84.7%	Score 14.4;	DB 10;	Length 25;
Best Local Similarity	93.8%	Pred. No. 1.7e+03;		
Matches 15: Conservative	0: Mismatches	1: Indels	0: Gaps	0: Gaps

Qy 2 CCACGCTTGCCGAG 17  
|||  
Db 1 CCACGCTTGCCCAAG 16

## RESULT 57

US-10-995-561-64090/c  
; Sequence 64090, Application US/10995561  
; Publication No. US20050272054A1  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele et al.  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF  
; TITLE OF INVENTION: DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001559  
; CURRENT APPLICATION NUMBER: US/10/995,561  
; CURRENT FILING DATE: 2004-11-24  
; NUMBER OF SEQ ID NOS: 85702  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 64090  
; LENGTH: 201  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-995-561-64090

Query Match 84.7%; Score 14.4; DB 10; Length 201;  
Best Local Similarity 93.8%; Pred. No. 1.3e+03;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCCACGCTTGGCCGAG 17  
|||||  
DB 158 CCCACGCTTGGCCGAG 143

## RESULT 58

US-09-803-719-429/c  
; Sequence 429, Application US/09803719  
; Publication No. US20030044783A1  
; GENERAL INFORMATION:  
; APPLICANT: Williams, Lewis T.  
; APPLICANT: Escobedo, Jaime  
; APPLICANT: Innis, Michael A.  
; APPLICANT: Garcia, Pablo Dominiguez  
; APPLICANT: Sudduth-Klinger, Julie  
; APPLICANT: Reinhard, Christoph  
; APPLICANT: Giese, Klaus  
; APPLICANT: Randazzo, Filippo  
; APPLICANT: Kennedy, Giulia C.  
; APPLICANT: Pot, David  
; APPLICANT: Kassam, Altaf  
; APPLICANT: Lamson, George  
; APPLICANT: Drmanac, Radoje  
; APPLICANT: Crkvenjakov, Radomir  
; APPLICANT: Dickson, Mark  
; APPLICANT: Drmanac, Snezana  
; APPLICANT: Labat, Ivan  
; APPLICANT: Leshkowitz, Dena  
; APPLICANT: Kita, David  
; APPLICANT: Garcia, Veronica  
; APPLICANT: Jones, Lee William  
; APPLICANT: Stache-Crain, Bitgit  
; TITLE OF INVENTION: Human Genes and Gene Products  
; FILE REFERENCE: 1624.002  
; CURRENT APPLICATION NUMBER: US/09/803,719  
; CURRENT FILING DATE: 2001-03-09  
; PRIOR APPLICATION NUMBER: 60/188,609  
; PRIOR FILING DATE: 2000-03-09  
; NUMBER OF SEQ ID NOS: 2396  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 429  
; LENGTH: 344  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(344)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-803-719-429

Query Match 84.7%; Score 14.4; DB 3; Length 344;  
Best Local Similarity 93.8%; Pred. No. 1.2e+03;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCCACGCTTGGCCGAG 17  
|||||  
DB 252 CCCACGCTTGGCCGAG 237

## RESULT 59

US-10-779-543-13699/c  
; Sequence 13699, Application US/10779543  
; Publication No. US20050227917A1  
; GENERAL INFORMATION:  
; APPLICANT: Williams et al  
; TITLE OF INVENTION: GENE PRODUCTS DIFFERENTIALLY EXPRESSED  
; TITLE OF INVENTION: IN CANCEROUS CELLS AND THEIR METHODS OF USE II  
; FILE REFERENCE: 2300-21302  
; CURRENT APPLICATION NUMBER: US/10/779,543  
; CURRENT FILING DATE: 2004-02-12  
; PRIOR APPLICATION NUMBER: 10/076,555  
; PRIOR FILING DATE: 2002-02-15  
; PRIOR APPLICATION NUMBER: 09/217,471  
; PRIOR FILING DATE: 1998-12-21  
; PRIOR APPLICATION NUMBER: 60/068,755  
; PRIOR FILING DATE: 1997-12-23  
; PRIOR APPLICATION NUMBER: 60/080,664  
; PRIOR FILING DATE: 1998-04-03  
; PRIOR APPLICATION NUMBER: 60/105,234  
; PRIOR FILING DATE: 1998-10-21  
; PRIOR APPLICATION NUMBER: 09/297,648  
; PRIOR FILING DATE: 2000-04-10  
; PRIOR APPLICATION NUMBER: PCT/US99/01619  
; PRIOR FILING DATE: 1999-01-28  
; PRIOR APPLICATION NUMBER: 60/072,910  
; PRIOR FILING DATE: 1998-01-28  
; PRIOR APPLICATION NUMBER: 60/075,954  
; PRIOR FILING DATE: 1998-02-24  
; PRIOR APPLICATION NUMBER: 60/080,114  
; PRIOR FILING DATE: 1998-03-31  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 23767  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13699  
; LENGTH: 344  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 344  
; OTHER INFORMATION: n = A,T,C or G  
US-10-779-543-13699

Query Match 84.7%; Score 14.4; DB 10; Length 344;  
Best Local Similarity 93.8%; Pred. No. 1.2e+03;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCCACGCTTGGCCGAG 17  
|||||  
DB 252 CCCACGCTTGGCCGAG 237

## RESULT 60

US-10-357-930-54041  
; Sequence 54041, Application US/10357930  
; Publication No. US20040259086A1  
; GENERAL INFORMATION:  
; APPLICANT: Schlegel, Robert  
; APPLICANT: Endege, Wilson  
; APPLICANT: Monahan, John  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR  
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF

; TITLE OF INVENTION: HUMAN PROSTATE CANCER  
; FILE REFERENCE: MRI-007BCN  
; CURRENT APPLICATION NUMBER: US/10/357,930  
; CURRENT FILING DATE: 2003-02-04  
; PRIOR APPLICATION NUMBER: 09/785,276  
; PRIOR FILING DATE: 2003-02-16  
; PRIOR APPLICATION NUMBER: 60/183,319  
; PRIOR FILING DATE: 2000-02-17  
; PRIOR APPLICATION NUMBER: 60/189,862  
; PRIOR FILING DATE: 2000-03-16  
; PRIOR APPLICATION NUMBER: 60/207,454  
; PRIOR FILING DATE: 2000-05-25  
; PRIOR APPLICATION NUMBER: 60/211,314  
; PRIOR FILING DATE: 2000-06-09  
; PRIOR APPLICATION NUMBER: 60/219,007  
; PRIOR FILING DATE: 2000-07-18  
; PRIOR APPLICATION NUMBER: 60/255,281  
; PRIOR FILING DATE: 2000-12-13  
; NUMBER OF SEQ ID NOS: 62232  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 54041  
; LENGTH: 570  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 28, 41, 42  
; OTHER INFORMATION: n = A,T,C or G  
US-10-357-930-54041

Query Match 84.7%; Score 14.4; DB 9; Length 570;  
Best Local Similarity 93.8%; Pred. No. 1.1e+03;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGCCGA 16  
Db 140 GCCCAGCGTTGCCGA 155

## RESULT 61

US-09-925-065A-939940/c  
; Sequence 939940, Application US/09925065A  
; Publication No. US20040181048A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; FILE REFERENCE: 108827,135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 939940  
; LENGTH: 653  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-939940

Query Match 84.7%; Score 14.4; DB 4; Length 653;  
Best Local Similarity 93.8%; Pred. No. 1.1e+03;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 CCCAGCGTTGCCGAG 17

Db 648 CCCAGCGTTGCCGAG 633

## RESULT 62

US-09-925-065A-939940/c  
; Sequence 939940, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; FILE REFERENCE: 108827,135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 939940  
; LENGTH: 653  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-939940

Query Match 84.7%; Score 14.4; DB 5; Length 653;  
Best Local Similarity 93.8%; Pred. No. 1.1e+03;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 CCCAGCGTTGCCGAG 17

Db 648 CCCAGCGTTGCCGAG 633

## RESULT 63

US-09-925-065A-63251/c  
; Sequence 63251, Application US/09925065A  
; Publication No. US20040181048A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; FILE REFERENCE: 108827,135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 63251  
; LENGTH: 834  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-63251

Query Match 84.7%; Score 14.4; DB 4; Length 834;  
Best Local Similarity 93.8%; Pred. No. 1e+03;

Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCCACGCTTGCCGAG 17  
Db 69 CCCACCTTGCCGAG 54

RESULT 64  
US-09-925-065A-63251/c  
; Sequence 63251, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; PRIOR FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 63251  
; LENGTH: 834  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-63251

Query Match 84.7%; Score 14.4; DB 5; Length 834;  
Best Local Similarity 93.8%; Pred. No. 1e+03;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCCACGCTTGCCGAG 17  
Db 69 CCCACCTTGCCGAG 54

RESULT 65  
US-10-301-480-164489/c  
; Sequence 164489, Application US/10301480  
; Publication No. US20060057564A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms  
; FILE REFERENCE: 108827.137  
; CURRENT APPLICATION NUMBER: US/10/301,480  
; PRIOR FILING DATE: 2002-11-21  
; PRIOR APPLICATION NUMBER: US 10/215,598  
; PRIOR FILING DATE: 2002-08-09  
; PRIOR APPLICATION NUMBER: US 60/311,695  
; PRIOR FILING DATE: 2001-08-10  
; NUMBER OF SEQ ID NOS: 1226818  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 164489  
; LENGTH: 834  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-10-301-480-164489

Query Match 84.7%; Score 14.4; DB 12; Length 834;  
Best Local Similarity 93.8%; Pred. No. 1e+03;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCCACGCTTGCCGAG 17

Db 69 CCCACCTTGCCGAG 54

RESULT 66  
US-10-301-480-777898/c  
; Sequence 777898, Application US/10301480  
; Publication No. US20060057564A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms  
; FILE REFERENCE: 108827.137  
; CURRENT APPLICATION NUMBER: US/10/301,480  
; PRIOR FILING DATE: 2002-11-21  
; PRIOR APPLICATION NUMBER: US 10/215,598  
; PRIOR FILING DATE: 2002-08-09  
; PRIOR APPLICATION NUMBER: US 60/311,695  
; PRIOR FILING DATE: 2001-08-10  
; NUMBER OF SEQ ID NOS: 1226818  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 777898  
; LENGTH: 834  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-10-301-480-777898

Query Match 84.7%; Score 14.4; DB 12; Length 834;  
Best Local Similarity 93.8%; Pred. No. 1e+03;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCCACGCTTGCCGAG 17  
Db 69 CCCACCTTGCCGAG 54

RESULT 67  
US-10-101-510-469/c  
; Sequence 469, Application US/10101510  
; Publication No. US20030148295A1  
; GENERAL INFORMATION:  
; APPLICANT: WAN, JACKSON  
; APPLICANT: WANG, YIXIN  
; TITLE OF INVENTION: EXPRESSION PROFILES AND METHODS OF USE  
; FILE REFERENCE: 15117.0012  
; CURRENT APPLICATION NUMBER: US/10/101,510  
; CURRENT FILING DATE: 2002-03-20  
; PRIOR APPLICATION NUMBER: 60/276,947  
; PRIOR FILING DATE: 2001-03-20  
; NUMBER OF SEQ ID NOS: 805  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 469  
; LENGTH: 1889  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-101-510-469

Query Match 84.7%; Score 14.4; DB 7; Length 1889;  
Best Local Similarity 93.8%; Pred. No. 9.4e+02;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCCACGCTTGCCGAG 17  
Db 18 CCCACGCTTGCCGAG 3

RESULT 68  
US-09-861-846-3/c  
; Sequence 3, Application US/09861846  
; Patent No. US20020110852A1  
; GENERAL INFORMATION:  
; APPLICANT: GUEGLER, Karl et al.  
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,

```

; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL001065
; CURRENT APPLICATION NUMBER: US/09/861,846
; CURRENT FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: 09/752,821
; PRIOR FILING DATE: 2001-01-03
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 13608
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(13608)
; OTHER INFORMATION: n = A,T,C or G
US-09-861-846-3

```

```

Query Match      84.7%; Score 14.4; DB 3; Length 13608;
Best Local Similarity 93.8%; Pred. No. 7.2e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      2 CCCAGCGTTGCCCGAG 17
Db      7742 CCCAGCGTTGGCTGAG 7727

```

## RESULT 69

```

US-10-250-463-3/c
; Sequence 3, Application US/10250463
; Publication No. US20040106775A1
; GENERAL INFORMATION:
; APPLICANT: PE CORPORATION (NY)
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL001065
; CURRENT APPLICATION NUMBER: US/10/250,463
; CURRENT FILING DATE: 2003-07-02
; PRIOR APPLICATION NUMBER: 09/752,821
; PRIOR FILING DATE: 2001-01-03
; PRIOR APPLICATION NUMBER: 09/861,846
; PRIOR FILING DATE: 2001-05-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 13608
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(13608)
; OTHER INFORMATION: n = A,T,C or G
US-10-250-463-3

```

```

Query Match      84.7%; Score 14.4; DB 8; Length 13608;
Best Local Similarity 93.8%; Pred. No. 7.2e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      2 CCCAGCGTTGCCCGAG 17
Db      7742 CCCAGCGTTGGCTGAG 7727

```

## RESULT 70

```

US-10-294-934-485/c
; Sequence 485, Application US/10294934
; Publication No. US20040038231A1
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Chumakov, Ilva
; APPLICANT: Bougueleret, Lydie

```

```

; APPLICANT: Cohen, Annick
; TITLE OF INVENTION: BIALLELIC MARKERS RELATED TO GENES INVOLVED IN DRUG METABOLISM
; FILE REFERENCE: 62.US4.DIV
; CURRENT APPLICATION NUMBER: US/10/294,934
; CURRENT FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US 09/671,317
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US 09/536,178
; PRIOR FILING DATE: 2000-03-23
; PRIOR APPLICATION NUMBER: PCT/IB00/00403
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: US 60/126,269
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/131,961
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 977
; SOFTWARE: Patent.pm
; SEQ ID NO 485
; LENGTH: 49312
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 5466..7466
; OTHER INFORMATION: 5'regulatory region
; FEATURE:
; NAME/KEY: exon
; LOCATION: 7467..7725
; OTHER INFORMATION: exon 1
; FEATURE:
; NAME/KEY: exon
; LOCATION: 20256..20355
; OTHER INFORMATION: exon 2
; FEATURE:
; NAME/KEY: exon
; LOCATION: 36905..36975
; OTHER INFORMATION: exon 3
; FEATURE:
; NAME/KEY: exon
; LOCATION: 45167..45248
; OTHER INFORMATION: exon 4
; FEATURE:
; NAME/KEY: exon
; LOCATION: 45728..45965
; OTHER INFORMATION: exon 5
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 45966..49312
; OTHER INFORMATION: 3'regulatory region
; FEATURE:
; NAME/KEY: allele
; LOCATION: 7564
; OTHER INFORMATION: 10-286-289 : polymorphic base G or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 7619
; OTHER INFORMATION: 10-286-345 : polymorphic base A or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 7649
; OTHER INFORMATION: 10-286-375 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 17258
; OTHER INFORMATION: 12-425-57 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 21590
; OTHER INFORMATION: 12-421-135 : insertion of T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 21595
; OTHER INFORMATION: 12-421-140 : polymorphic base A or G

```



```
;
; NAME/KEY: allele
; LOCATION: 36971
; OTHER INFORMATION: 10-523-232 : polymorphic base C or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 45214
; OTHER INFORMATION: 10-289-201 : polymorphic base C or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 45741
; OTHER INFORMATION: 10-290-37 : polymorphic base C or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 46029
; OTHER INFORMATION: 10-290-326 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 46032
; OTHER INFORMATION: 10-290-328 : deletion of G
; FEATURE:
; NAME/KEY: primer bind
; LOCATION: 7276..7294
; OTHER INFORMATION: 10-286.pu
; FEATURE:
; NAME/KEY: primer bind
; LOCATION: 7676..7694
; OTHER INFORMATION: 10-286.rp complement
; FEATURE:
; NAME/KEY: primer bind
; LOCATION: 16839..16856
; OTHER INFORMATION: 12-425.rp
; FEATURE:
; NAME/KEY: primer bind
; LOCATION: 17297..17314
; OTHER INFORMATION: 12-425.pu complement
; FEATURE:
; NAME/KEY: primer bind
; LOCATION: 21456..21474
; OTHER INFORMATION: 12-421.pu
; FEATURE:
; NAME/KEY: primer bind
; LOCATION: 21886..21906
; OTHER INFORMATION: 12-421.rp complement
; FEATURE:
; NAME/KEY: primer bind
; LOCATION: 36740..36758
; OTHER INFORMATION: 10-523.pu
; FEATURE:
; NAME/KEY: primer bind
; LOCATION: 36997..37015
; OTHER INFORMATION: 10-523.rp complement
; FEATURE:
; NAME/KEY: primer bind
; LOCATION: 45020..45037
; OTHER INFORMATION: 10-289.pu
; FEATURE:
; NAME/KEY: primer bind
; LOCATION: 45413..45432
; OTHER INFORMATION: 10-289.rp complement
; FEATURE:
; NAME/KEY: primer bind
; LOCATION: 45705..45724
; OTHER INFORMATION: 10-290.pu
; FEATURE:
; NAME/KEY: primer bind
; LOCATION: 46104..46123
; OTHER INFORMATION: 10-290.rp complement
; FEATURE:
; NAME/KEY: primer bind
; LOCATION: 7545..7563
; OTHER INFORMATION: 10-286-289.mis
; FEATURE:
```

```
;
; NAME/KEY: primer bind
; LOCATION: 7565..7583
; OTHER INFORMATION: 10-286-289.mis complement
; FEATURE:
; NAME/KEY: primer bind
; LOCATION: 7600..7618
; OTHER INFORMATION: 10-286-345.mis
; FEATURE:
; NAME/KEY: primer bind
; LOCATION: 7620..7638
; OTHER INFORMATION: 10-286-345.mis complement
; FEATURE:
; NAME/KEY: primer bind
; LOCATION: 7630..7648
; OTHER INFORMATION: 10-286-375.mis
; FEATURE:
; NAME/KEY: primer bind
; LOCATION: 7650..7668
; OTHER INFORMATION: 10-286-375.mis complement
; FEATURE:
; NAME/KEY: primer bind
; LOCATION: 17239..17257
; OTHER INFORMATION: 12-425-57.mis
; FEATURE:
; NAME/KEY: primer bind
; LOCATION: 17259..17277
; OTHER INFORMATION: 12-425-57.mis complement
; FEATURE:
; NAME/KEY: primer bind
; LOCATION: 21576..21594
; OTHER INFORMATION: 12-421-140.mis
; FEATURE:
; NAME/KEY: primer bind
; LOCATION: 21596..21614
; OTHER INFORMATION: 12-421-140.mis complement
; FEATURE:
; NAME/KEY: primer bind
; LOCATION: 36952..36970
; OTHER INFORMATION: 10-523-232.mis
; FEATURE:
; NAME/KEY: primer bind
; LOCATION: 36972..36990
; OTHER INFORMATION: 10-523-232.mis complement
; FEATURE:
; NAME/KEY: primer bind
; LOCATION: 45195..45213
; OTHER INFORMATION: 10-289-201.mis
; FEATURE:
```

```
Query Match      84.7%; Score 14.4; DB 8; Length 49312;
Best Local Similarity 93.8%; Pred. No. 6.1e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 2 CCCACGGTTGCCCGAG 17
Db 12970 CCCACTCTGGCCGAG 12955
|||||
```

```
RESULT 71
US-11-204-311-485/c
; Sequence 485, Application US/11204311
; Publication No. US20060040304A1
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Chumakov, Ilya
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Cohen, Annick
; TITLE OF INVENTION: BIALLELIC MARKERS RELATED TO GENES INVOLVED IN DRUG METABOLISM
; FILE REFERENCE: 62 US4 DIV
; CURRENT APPLICATION NUMBER: US/11/204,311
; CURRENT FILING DATE: 2005-08-15
; PRIOR APPLICATION NUMBER: US/10/294,934
; PRIOR FILING DATE: 2000-09-27
```

```

; PRIOR APPLICATION NUMBER: US 09/671,317
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US 09/536,178
; PRIOR FILING DATE: 2000-03-23
; PRIOR APPLICATION NUMBER: PCT/IB00/00403
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: US 60/126,269
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/131,961
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 977
; SOFTWARE: Patent.pm
; SEQ ID NO 485
; LENGTH: 49312
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 5466..7466
; OTHER INFORMATION: 5'regulatory region
; FEATURE:
; NAME/KEY: exon
; LOCATION: 7467..7725
; OTHER INFORMATION: exon 1
; FEATURE:
; NAME/KEY: exon
; LOCATION: 20256..20355
; OTHER INFORMATION: exon 2
; FEATURE:
; NAME/KEY: exon
; LOCATION: 36905..36975
; OTHER INFORMATION: exon 3
; FEATURE:
; NAME/KEY: exon
; LOCATION: 45167..45248
; OTHER INFORMATION: exon 4
; FEATURE:
; NAME/KEY: exon
; LOCATION: 45728..45965
; OTHER INFORMATION: exon 5
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 45966..49312
; OTHER INFORMATION: 3'regulatory region
; FEATURE:
; NAME/KEY: allele
; LOCATION: 7564
; OTHER INFORMATION: 10-286-289 : polymorphic base G or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 7619
; OTHER INFORMATION: 10-286-345 : polymorphic base A or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 7649
; OTHER INFORMATION: 10-286-375 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 17258
; OTHER INFORMATION: 12-425-57 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 21590
; OTHER INFORMATION: 12-421-135 : insertion of T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 21595
; OTHER INFORMATION: 12-421-140 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 36971
; OTHER INFORMATION: 10-523-232 : polymorphic base C or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 45214
; OTHER INFORMATION: 10-289-201 : polymorphic base C or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 45741
; OTHER INFORMATION: 10-290-37 : polymorphic base C or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 46029
; OTHER INFORMATION: 10-290-326 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 46032
; OTHER INFORMATION: 10-290-328 : deletion of G
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 7276..7294
; OTHER INFORMATION: 10-286.pu
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 7676..7694
; OTHER INFORMATION: 10-286.rp complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 16839..16856
; OTHER INFORMATION: 12-425.rp
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 17297..17314
; OTHER INFORMATION: 12-425.pu complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 21456..21474
; OTHER INFORMATION: 12-421.pu
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 21886..21906
; OTHER INFORMATION: 12-421.rp complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 36740..36758
; OTHER INFORMATION: 10-523.pu
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 36997..37015
; OTHER INFORMATION: 10-523.rp complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 45020..45037
; OTHER INFORMATION: 10-289.pu
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 45413..45432
; OTHER INFORMATION: 10-289.rp complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 45705..45724
; OTHER INFORMATION: 10-290.pu
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 46104..46123
; OTHER INFORMATION: 10-290.rp complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 7545..7563
; OTHER INFORMATION: 10-286-289.mis
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 7565..7583
; OTHER INFORMATION: 10-286-289.mis complement
; FEATURE:
; NAME/KEY: primer_bind

```



```
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 63916
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-63916

Query Match      82.4%; Score 14; DB 10; Length 201;
Best Local Similarity 87.5%; Pred. No. 2.1e+03;
Matches 14; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      2 CCCACGCTTGCCCGAG 17
Db      101 MCCAGGCTTGCCCGAG 86

RESULT 76
US-10-301-480-242988/c
; Sequence 242988, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 242988
; LENGTH: 429
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-242988

Query Match      82.4%; Score 14; DB 12; Length 429;
Best Local Similarity 87.5%; Pred. No. 1.9e+03;
Matches 14; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      2 CCCACGCTTGCCCGAG 17
Db      280 CCCACGCTTGCCCGAG 265

RESULT 77
US-10-301-480-856397/c
; Sequence 856397, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 856397
; LENGTH: 429
; TYPE: DNA
; ORGANISM: Homo sapien
```

```
US-10-301-480-856397

Query Match      82.4%; Score 14; DB 12; Length 429;
Best Local Similarity 87.5%; Pred. No. 1.9e+03;
Matches 14; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      2 CCCACGCTTGCCCGAG 17
Db      280 CCCACGCTTGCCCGAG 265

RESULT 78
US-09-925-065A-148858/c
; Sequence 148858, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 148858
; LENGTH: 440
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-148858

Query Match      82.4%; Score 14; DB 4; Length 440;
Best Local Similarity 87.5%; Pred. No. 1.9e+03;
Matches 14; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      2 CCCACGCTTGCCCGAG 17
Db      279 CCCACGCTTGCCCGAG 264

RESULT 79
US-09-925-065A-148858/c
; Sequence 148858, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 148858
```

LENGTH: 440  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-925-065A-148858

Query Match 82.4%; Score 14; DB 5; Length 440;  
Best Local Similarity 87.5%; Pred. No. 1.9e+03;  
Matches 14; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 CCACGCTTGGCCGAG 17  
|||||:|||||  
Db 279 CCCACGCTGGCCGAG 264

## RESULT 80

US-09-918-995-28686/C  
Sequence 28686, Application US/09918995  
Publication No. US20030073623A1  
GENERAL INFORMATION:  
APPLICANT: Hyseq, Inc.  
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
FILE REFERENCE: 20411-756  
CURRENT APPLICATION NUMBER: US/09/918,995  
CURRENT FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: US/09/235,076  
PRIOR FILING DATE: 1999-01-20  
NUMBER OF SEQ ID NOS: 38054  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 28686  
LENGTH: 458

TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(458)  
OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-28686

Query Match 82.4%; Score 14; DB 3; Length 458;  
Best Local Similarity 100.0%; Pred. No. 1.8e+03;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 CCACGCTTGGCCGA 16  
|||||:|||||  
Db 358 CCACGCTTGGCCGA 345

## RESULT 81

US-10-674-124A-19328  
Sequence 19328, Application US/10674124A  
Publication No. US20040197797A1  
GENERAL INFORMATION:  
APPLICANT: INOKO, Hidetoshi  
APPLICANT: TAMIYA, Gen  
TITLE OF INVENTION: GENE MAPPING METHOD USING MICROSATELLITE  
TITLE OF INVENTION: GENETIC POLYMORPHISM MARKERS  
FILE REFERENCE: ORIN-003CIP  
CURRENT APPLICATION NUMBER: US/10/674,124A  
CURRENT FILING DATE: 2003-09-26  
PRIOR APPLICATION NUMBER: 10/257,511  
PRIOR FILING DATE: 2003-03-07  
PRIOR APPLICATION NUMBER: PCT/JP00/07621  
PRIOR FILING DATE: 2000-10-30  
PRIOR APPLICATION NUMBER: JP2000-112699  
PRIOR FILING DATE: 2000-04-13  
PRIOR APPLICATION NUMBER: JP2002-327516  
PRIOR FILING DATE: 2002-09-28  
PRIOR APPLICATION NUMBER: JP2002-383869  
PRIOR FILING DATE: 2002-12-09  
NUMBER OF SEQ ID NOS: 27110  
SEQ ID NO 19328  
LENGTH: 465

TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: AC073575.5\_144468  
FEATURE:  
OTHER INFORMATION: Located on chromosome 12  
FEATURE:  
OTHER INFORMATION: Distance between a terminus base of telomere on  
OTHER INFORMATION: chromosomal short arm and 5'-terminus of this base  
OTHER INFORMATION: sequence : 118577396  
FEATURE:  
OTHER INFORMATION: Distance between 3'-terminus of neighbour sequence of  
OTHER INFORMATION: sequence listing upward to telomere on chromosomal short arm and  
OTHER INFORMATION: 5'-terminus of this base sequence : 99235  
US-10-674-124A-19328

Query Match 82.4%; Score 14; DB 9; Length 465;  
Best Local Similarity 100.0%; Pred. No. 1.8e+03;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 CCACGCTTGGCCGA 16  
|||||:|||||  
Db 401 CCACGCTTGGCCGA 414

## RESULT 82

US-09-925-065A-82285  
Sequence 82285, Application US/09925065A  
Publication No. US20040181048A1  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single  
TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome  
FILE REFERENCE: 108827.135  
CURRENT APPLICATION NUMBER: US/09/925,065A  
CURRENT FILING DATE: 2001-08-08  
PRIOR APPLICATION NUMBER: US 60/243,096  
PRIOR FILING DATE: 2000-10-24  
PRIOR APPLICATION NUMBER: US 60/252,147  
PRIOR FILING DATE: 2000-11-20  
PRIOR APPLICATION NUMBER: US 60/250,092  
PRIOR FILING DATE: 2000-11-30  
PRIOR APPLICATION NUMBER: US 60/261,766  
PRIOR FILING DATE: 2001-01-16  
PRIOR APPLICATION NUMBER: US 60/289,846  
PRIOR FILING DATE: 2001-05-09  
NUMBER OF SEQ ID NOS: 957086  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 82285  
LENGTH: 518

TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-925-065A-82285

Query Match 82.4%; Score 14; DB 4; Length 518;  
Best Local Similarity 87.5%; Pred. No. 1.8e+03;  
Matches 14; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCACGCTTGGCCGA 16  
|||||:|||||  
Db 237 GCCCACGCTTGGCCGA 252

## RESULT 83

US-09-925-065A-82285  
Sequence 82285, Application US/09925065A  
Publication No. US20050228172A9  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single  
TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome  
FILE REFERENCE: 108827.135  
CURRENT APPLICATION NUMBER: US/09/925,065A

```
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 82285
; LENGTH: 518
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-82285

Query Match      82.4%; Score 14; DB 5; Length 518;
Best Local Similarity 87.5%; Pred. No. 1.8e+03;
Matches 14; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      1 GCCCAGCGTTGGCCGA 16
Db      237 GCCCAGCGTTGGCCRA 252

RESULT 84
US-10-301-480-183525
; Sequence 183525, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms
; TITLE OF INVENTION: in the Human Genome
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 183525
; LENGTH: 518
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-183525

Query Match      82.4%; Score 14; DB 12; Length 518;
Best Local Similarity 87.5%; Pred. No. 1.8e+03;
Matches 14; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      1 GCCCAGCGTTGGCCGA 16
Db      237 GCCCAGCGTTGGCCRA 252

RESULT 85
US-10-301-480-796934
; Sequence 796934, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms
; TITLE OF INVENTION: in the Human Genome
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
```

```
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 796934
; LENGTH: 518
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-796934

Query Match      82.4%; Score 14; DB 12; Length 518;
Best Local Similarity 87.5%; Pred. No. 1.8e+03;
Matches 14; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      1 GCCCAGCGTTGGCCGA 16
Db      237 GCCCAGCGTTGGCCRA 252

RESULT 86
US-09-925-065A-805681/c
; Sequence 805681, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 805681
; LENGTH: 614
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-805681

Query Match      82.4%; Score 14; DB 4; Length 614;
Best Local Similarity 100.0%; Pred. No. 1.8e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      3 CCACGCTTGGCCGA 16
Db      435 CCACGCTTGGCCGA 422

RESULT 87
US-09-925-065A-805681/c
; Sequence 805681, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
```

; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 805681  
; LENGTH: 614  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-805681

Query Match 82.4%; Score 14; DB 5; Length 614;  
Best Local Similarity 100.0%; Pred. No. 1.8e+03;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 CCACGCTTGGCCGA 16  
||| ||||| |||||  
Db 435 CCACGCTTGGCCGA 422

RESULT 88  
US-09-925-065A-14  
; Sequence 14, Application US/09925065A  
; Publication No. US20040181048A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14  
; LENGTH: 683  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-14

Query Match 82.4%; Score 14; DB 4; Length 683;  
Best Local Similarity 100.0%; Pred. No. 1.7e+03;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 CCACGCTTGGCCGA 16  
||| ||||| |||||  
Db 99 CCACGCTTGGCCGA 112

RESULT 89  
US-09-925-065A-15  
; Sequence 15, Application US/09925065A  
; Publication No. US20040181048A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24

; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 15  
; LENGTH: 683  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-15

Query Match 82.4%; Score 14; DB 4; Length 683;  
Best Local Similarity 100.0%; Pred. No. 1.7e+03;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 CCACGCTTGGCCGA 16  
||| ||||| |||||  
Db 99 CCACGCTTGGCCGA 112

RESULT 90  
US-09-925-065A-14  
; Sequence 14, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14  
; LENGTH: 683  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-14

Query Match 82.4%; Score 14; DB 5; Length 683;  
Best Local Similarity 100.0%; Pred. No. 1.7e+03;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 CCACGCTTGGCCGA 16  
||| ||||| |||||  
Db 99 CCACGCTTGGCCGA 112

RESULT 91  
US-09-925-065A-15  
; Sequence 15, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A

; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 15  
; LENGTH: 683  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-15

Query Match 82.4%; Score 14; DB 5; Length 683;  
Best Local Similarity 100.0%; Pred. No. 1.7e+03;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 CCACGCTTGCCGA 16  
|||||  
Db 99 CCACGCTTGCCGA 112

## RESULT 92

US-10-301-480-101250  
; Sequence 101250, Application US/10301480  
; Publication No. US20060057564A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms  
; FILE REFERENCE: 108827.137  
; CURRENT APPLICATION NUMBER: US/10/301,480  
; CURRENT FILING DATE: 2002-11-21  
; PRIOR APPLICATION NUMBER: US 10/215,598  
; PRIOR FILING DATE: 2002-08-09  
; PRIOR APPLICATION NUMBER: US 60/311,695  
; PRIOR FILING DATE: 2001-08-10  
; NUMBER OF SEQ ID NOS: 1226818  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 101250  
; LENGTH: 683  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-10-301-480-101250

Query Match 82.4%; Score 14; DB 12; Length 683;  
Best Local Similarity 100.0%; Pred. No. 1.7e+03;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 CCACGCTTGCCGA 16  
|||||  
Db 99 CCACGCTTGCCGA 112

## RESULT 93

US-10-301-480-101251  
; Sequence 101251, Application US/10301480  
; Publication No. US20060057564A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms  
; FILE REFERENCE: 108827.137  
; CURRENT APPLICATION NUMBER: US/10/301,480  
; CURRENT FILING DATE: 2002-11-21  
; PRIOR APPLICATION NUMBER: US 10/215,598  
; PRIOR FILING DATE: 2002-08-09

; PRIOR APPLICATION NUMBER: US 60/311,695  
; PRIOR FILING DATE: 2001-08-10  
; NUMBER OF SEQ ID NOS: 1226818  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 101251  
; LENGTH: 683  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-10-301-480-101251

Query Match 82.4%; Score 14; DB 12; Length 683;  
Best Local Similarity 100.0%; Pred. No. 1.7e+03;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 CCACGCTTGCCGA 16  
|||||  
Db 99 CCACGCTTGCCGA 112

## RESULT 94

US-10-301-480-714659  
; Sequence 714659, Application US/10301480  
; Publication No. US20060057564A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms  
; FILE REFERENCE: 108827.137  
; CURRENT APPLICATION NUMBER: US/10/301,480  
; CURRENT FILING DATE: 2002-11-21  
; PRIOR APPLICATION NUMBER: US 10/215,598  
; PRIOR FILING DATE: 2002-08-09  
; PRIOR APPLICATION NUMBER: US 60/311,695  
; PRIOR FILING DATE: 2001-08-10  
; NUMBER OF SEQ ID NOS: 1226818  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 714659  
; LENGTH: 683  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-10-301-480-714659

Query Match 82.4%; Score 14; DB 12; Length 683;  
Best Local Similarity 100.0%; Pred. No. 1.7e+03;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 CCACGCTTGCCGA 16  
|||||  
Db 99 CCACGCTTGCCGA 112

## RESULT 95

US-10-301-480-714660  
; Sequence 714660, Application US/10301480  
; Publication No. US20060057564A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms  
; FILE REFERENCE: 108827.137  
; CURRENT APPLICATION NUMBER: US/10/301,480  
; CURRENT FILING DATE: 2002-11-21  
; PRIOR APPLICATION NUMBER: US 10/215,598  
; PRIOR FILING DATE: 2002-08-09  
; PRIOR APPLICATION NUMBER: US 60/311,695  
; PRIOR FILING DATE: 2001-08-10  
; NUMBER OF SEQ ID NOS: 1226818  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 714660  
; LENGTH: 683  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-10-301-480-714660



```
Query Match      82.4%; Score 14; DB 12; Length 683;
Best Local Similarity 100.0%; Pred. No. 1.7e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 CCACGCTTGGCCGA 16
Db 99 CCACGCTTGGCCGA 112

RESULT 96
US-10-301-480-598304/c
; Sequence 598304, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 598304
; LENGTH: 911
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-598304

Query Match      82.4%; Score 14; DB 12; Length 911;
Best Local Similarity 100.0%; Pred. No. 1.7e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 CCACGCTTGGCCGA 16
Db 212 CCACGCTTGGCCGA 199

RESULT 97
US-10-301-480-1211713/c
; Sequence 1211713, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1211713
; LENGTH: 911
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-1211713

Query Match      82.4%; Score 14; DB 12; Length 911;
Best Local Similarity 100.0%; Pred. No. 1.7e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 CCACGCTTGGCCGA 16
Db 212 CCACGCTTGGCCGA 199
```

```
RESULT 98
US-10-115-831-63/c
; Sequence 63, Application US/10115831
; Publication No. US20030219743A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Ren, Feiyun
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20030219743A1el Nucleic Acids and
; FILE REFERENCE: 792CIP2ADIV
; CURRENT APPLICATION NUMBER: US/10/115,831
; CURRENT FILING DATE: 2002-04-02
; PRIOR APPLICATION NUMBER: 09/667,298
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: 09/577,408
; PRIOR FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 178
; SOFTWARE: pt_FL_genes Version 2.0
; SEQ ID NO 63
; LENGTH: 1181
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (116)..(964)
US-10-115-831-63

Query Match      82.4%; Score 14; DB 7; Length 1181;
Best Local Similarity 100.0%; Pred. No. 1.6e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 CCCACGCTTGGCCG 15
Db 167 CCCACGCTTGGCCG 154

RESULT 99
US-11-096-568A-11881/c
; Sequence 11881, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nickolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 11881
; LENGTH: 2475
; TYPE: DNA
; ORGANISM: Triticum aestivum
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(2475)
; OTHER INFORMATION: Ceres Seq. ID no. 13658778
US-11-096-568A-11881

Query Match      82.4%; Score 14; DB 16; Length 2475;
Best Local Similarity 100.0%; Pred. No. 1.5e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCACGCTTGGCC 14
Db 1894 GCCCACGCTTGGCC 1881

RESULT 100
US-10-995-561-13368/c
```

```
; Sequence 13368, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: C0001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13368
; LENGTH: 70513
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-13368
```

```
Query Match      82.4%; Score 14; DB 10; Length 70513;
Best Local Similarity 87.5%; Pred. No. 9.4e+02;
Matches 14; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      2 CCCACGCTTGGCCGAG 17
Db      43202 MCCAGGCTTGGCCGAG 43187
```

```
Search completed: June 10, 2006, 20:52:22
Job time : 690.1 secs
```

GenCore version 5.1.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2006, 15:34:22 ; Search time 8.21667 Seconds  
(without alignments)  
262.600 Million cell updates/sec

Title: US-10-600-816-31

Perfect score: 17

Sequence: 1 gccacgcttgccgag 17

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 253354 seqs, 63461778 residues

Total number of hits satisfying chosen parameters: 506708

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

Database :

Published Applications NA New:\*

- 1: /EMC\_Celerra\_SIDS3/ptodata/1/pubpna/US09\_NEW\_PUB\_seq.\*
- 2: /EMC\_Celerra\_SIDS3/ptodata/1/pubpna/US06\_NEW\_PUB\_seq.\*
- 3: /EMC\_Celerra\_SIDS3/ptodata/1/pubpna/US07\_NEW\_PUB\_seq.\*
- 4: /EMC\_Celerra\_SIDS3/ptodata/1/pubpna/US08\_NEW\_PUB\_seq.\*
- 5: /EMC\_Celerra\_SIDS3/ptodata/1/pubpna/US10\_NEW\_PUB\_seq.\*
- 6: /EMC\_Celerra\_SIDS3/ptodata/1/pubpna/US11\_NEW\_PUB\_seq.\*
- 7: /EMC\_Celerra\_SIDS3/ptodata/1/pubpna/US12\_NEW\_PUB\_seq.\*
- 8: /EMC\_Celerra\_SIDS3/ptodata/1/pubpna/US13\_NEW\_PUB\_seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	13.8	81.2	2165	6	US-10-511-937-485
2	13.8	81.2	5382	1	US-09-484-331-21
3	13.8	81.2	7873	6	US-10-485-397-7
4	13.8	81.2	7943	6	US-10-485-397-8
5	13.8	81.2	9737	1	US-09-484-331-22
6	13.8	81.2	9737	1	US-09-484-331-23
7	13.8	81.2	9737	1	US-09-484-331-28
8	13.8	81.2	9871	1	US-09-484-331-24
9	13.8	81.2	10060	1	US-09-484-331-25
c 10	13.4	78.8	806	6	US-10-953-349-15149
c 11	13.4	78.8	2019	6	US-10-953-349-35663
c 12	13.4	78.8	2175	7	US-11-293-697-559
c 13	13.4	78.8	2811	6	US-10-511-937-648
14	13.4	78.8	3048	7	US-11-293-697-950
15	12.8	75.3	39	7	US-11-169-140-80
16	12.8	75.3	39	7	US-11-169-140-84
17	12.8	75.3	39	7	US-11-319-606-12
c 18	12.8	75.3	500	7	US-11-249-305-1
19	12.8	75.3	500	7	US-11-249-305-2
c 20	12.8	75.3	500	7	US-11-249-305-3
21	12.8	75.3	500	7	US-11-249-305-4
c 22	12.8	75.3	693	7	US-11-217-529-82759
c 23	12.8	75.3	726	6	US-10-953-349-20941
c 24	12.8	75.3	777	7	US-11-217-529-82116
25	12.8	75.3	1086	7	US-11-217-529-1630

c 26	12.8	75.3	1093	6	US-10-953-349-25892	Sequence 25892, A
c 27	12.8	75.3	1124	6	US-10-953-349-33000	Sequence 33000, A
c 28	12.8	75.3	1541	6	US-10-953-349-35987	Sequence 35987, A
c 29	12.8	75.3	1824	6	US-10-953-349-33688	Sequence 33688, A
c 30	12.8	75.3	2525	7	US-11-293-697-1055	Sequence 1055, Ap
c 31	12.8	75.3	2557	7	US-11-293-697-1132	Sequence 1132, Ap
c 32	12.8	75.3	2593	7	US-11-293-697-845	Sequence 845, App
c 33	12.8	75.3	2641	7	US-11-293-697-1461	Sequence 1461, Ap
c 34	12.8	75.3	2825	7	US-11-293-697-1497	Sequence 1497, Ap
c 35	12.8	75.3	3039	7	US-11-293-697-885	Sequence 885, App
c 36	12.8	75.3	3095	7	US-11-293-697-291	Sequence 291, App
c 37	12.8	75.3	3142	7	US-11-254-185-1	Sequence 1, Appli
c 38	12.8	75.3	3142	7	US-11-253-869-1	Sequence 1, Appli
c 39	12.8	75.3	3693	7	US-11-293-697-510	Sequence 510, App
c 40	12.8	75.3	5202	7	US-11-121-154-158	Sequence 158, App
c 41	12.8	75.3	9426	7	US-11-217-529-1662	Sequence 1662, Ap
c 42	12.8	75.3	10905	6	US-10-505-928-568	Sequence 568, App
c 43	12.8	75.3	27419	6	US-10-857-260-34	Sequence 34, Appl
c 44	12.8	75.3	128361	6	US-10-505-928-151	Sequence 151, App
c 45	12.8	75.3	138941	6	US-10-489-730-10	GENERAL INFORMATI
c 46	12.4	72.9	39	7	US-11-169-140-82	Sequence 82, Appl
c 47	12.4	72.9	39	7	US-11-169-140-87	Sequence 87, Appl
c 48	12.4	72.9	39	7	US-11-169-140-95	Sequence 95, Appl
c 49	12.4	72.9	506	6	US-10-953-349-26326	Sequence 26326, A
c 50	12.4	72.9	1035	6	US-10-953-349-37335	Sequence 37335, A
c 51	12.4	72.9	1242	7	US-11-217-529-272	Sequence 272, App
c 52	12.4	72.9	1305	6	US-10-471-571A-2011	Sequence 2011, Ap
c 53	12.4	72.9	1423	6	US-10-953-349-35503	Sequence 35503, A
c 54	12.4	72.9	1429	6	US-10-953-349-35104	Sequence 35104, A
c 55	12.4	72.9	1685	7	US-11-293-697-3	Sequence 3, Appli
c 56	12.4	72.9	2175	7	US-11-293-697-559	Sequence 559, App
c 57	12.4	72.9	2299	7	US-11-293-697-322	Sequence 322, App
c 58	12.4	72.9	2728	7	US-11-293-697-1089	Sequence 1089, Ap
c 59	12.4	72.9	2750	7	US-11-293-697-1882	Sequence 1882, Ap
c 60	12.4	72.9	2882	6	US-10-511-937-576	Sequence 576, App
c 61	12.4	72.9	3093	7	US-11-217-529-3509	Sequence 3509, Ap
c 62	12.4	72.9	16032	6	US-10-527-552-5	Sequence 5, Appli
c 63	12.4	72.9	118899	7	US-11-189-279-64	Sequence 64, Appl
c 64	12.2	71.8	522	7	US-11-259-950-11	Sequence 11, Appl
c 65	12.2	71.8	522	7	US-11-217-997-9	Sequence 9, Appli
c 66	12.2	71.8	522	7	US-11-217-997-35	Sequence 35, Appl
c 67	12.2	71.8	540	7	US-11-330-650-2	Sequence 2, Appli
c 68	12.2	71.8	562	6	US-10-488-613-1465	Sequence 1465, Ap
c 69	12.2	71.8	619	7	US-11-259-950-5	Sequence 5, Appli
c 70	12.2	71.8	633	6	US-10-953-349-14927	Sequence 14927, A
c 71	12.2	71.8	654	6	US-11-253-950-3	Sequence 3, Appli
c 72	12.2	71.8	655	6	US-10-488-619-2847	Sequence 2847, Ap
c 73	12.2	71.8	681	7	US-11-217-529-566	Sequence 566, App
c 74	12.2	71.8	762	7	US-11-259-950-1	Sequence 1, Appli
c 75	12.2	71.8	798	6	US-10-953-349-28486	Sequence 28486, A
c 76	12.2	71.8	843	6	US-10-196-749-487	Sequence 487, App
c 77	12.2	71.8	868	6	US-10-953-349-20319	Sequence 20319, A
c 78	12.2	71.8	877	7	US-11-217-997-7	Sequence 7, Appli
c 79	12.2	71.8	945	6	US-10-953-349-28667	Sequence 28667, A
c 80	12.2	71.8	949	7	US-11-256-428-62	Sequence 62, Appl
c 81	12.2	71.8	993	6	US-10-953-349-27232	Sequence 27232, A
c 82	12.2	71.8	1038	6	US-10-953-349-3387	Sequence 3387, Ap
c 83	12.2	71.8	1233	6	US-10-953-349-35759	Sequence 35759, A
c 84	12.2	71.8	1233	6	US-11-254-252-1	Sequence 1, Appli
c 85	12.2	71.8	1260	7	US-11-259-950-7	Sequence 7, Appli
c 86	12.2	71.8	1275	7	US-11-259-950-77	Sequence 77, Appl
c 87	12.2	71.8	1282	7	US-11-259-950-9	Sequence 9, Appli
c 88	12.2	71.8	1295	6	US-10-953-349-35397	Sequence 35397, A
c 89	12.2	71.8	1378	6	US-10-953-349-27665	Sequence 27665, A
c 90	12.2	71.8	1409	6	US-10-953-349-28083	Sequence 28083, A
c 91	12.2	71.8	1605	6	US-10-953-349-35361	Sequence 35361, A
c 92	12.2	71.8	1646	6	US-10-505-928-653	Sequence 653, App
c 93	12.2	71.8	1736	7	US-11-293-697-981	Sequence 981, App
c 94	12.2	71.8	1859	6	US-10-953-349-34645	Sequence 34645, A
c 95	12.2	71.8	1908	7	US-11-254-252-13	Sequence 13, Appl
c 96	12.2	71.8	1941	7	US-11-217-529-79927	Sequence 79927, A
c 97	12.2	71.8	2128	7	US-11-293-697-2383	Sequence 2383, Ap
c 98	12.2	71.8	2182	7	US-11-293-697-1876	Sequence 1876, Ap















```
c 975 10.6 62.4 1408 6 US-10-953-349-30818
c 976 10.6 62.4 1411 6 US-10-953-349-38703
c 977 10.6 62.4 1437 7 US-11-217-529-78974
c 978 10.6 62.4 1439 6 US-10-953-349-32310
c 979 10.6 62.4 1446 6 US-10-953-349-35850
c 980 10.6 62.4 1447 7 US-11-243-296A-4
c 981 10.6 62.4 1449 7 US-11-217-529-2200
c 982 10.6 62.4 1451 6 US-10-953-349-31780
c 983 10.6 62.4 1464 6 US-10-953-349-29604
c 984 10.6 62.4 1466 6 US-10-505-928-458
c 985 10.6 62.4 1466 6 US-10-505-928-458
c 986 10.6 62.4 1485 6 US-10-953-349-24486
c 987 10.6 62.4 1486 6 US-10-953-349-22309
c 988 10.6 62.4 1486 6 US-10-953-349-38908
c 989 10.6 62.4 1504 6 US-10-953-349-35108
c 990 10.6 62.4 1517 6 US-10-953-349-35879
c 991 10.6 62.4 1529 6 US-10-953-349-33663
c 992 10.6 62.4 1531 6 US-10-953-349-34579
c 993 10.6 62.4 1533 6 US-10-953-349-21987
c 994 10.6 62.4 1541 6 US-10-953-349-12807
c 995 10.6 62.4 1555 6 US-10-953-349-34789
c 996 10.6 62.4 1556 6 US-10-953-349-30645
c 997 10.6 62.4 1568 6 US-10-953-349-24028
c 998 10.6 62.4 1571 6 US-10-713-648A-39
c 999 10.6 62.4 1576 6 US-10-953-349-7463
c1000 10.6 62.4 1579 7 US-11-293-697-1142

Sequence 30818, A
Sequence 38703, A
Sequence 78974, A
Sequence 32310, A
Sequence 35850, A
Sequence 4, Appl
Sequence 2200, Ap
Sequence 31780, A
Sequence 29604, A
Sequence 458, App
Sequence 458, App
Sequence 24486, A
Sequence 22309, A
Sequence 38908, A
Sequence 35108, A
Sequence 35879, A
Sequence 33663, A
Sequence 34579, A
Sequence 21987, A
Sequence 12807, A
Sequence 34789, A
Sequence 30645, A
Sequence 24028, A
Sequence 39, Appl
Sequence 7463, Ap
Sequence 1142, Ap

ALIGNMENTS

US-10-511-937-485/c
; Sequence 485, Application US/10511937
; Publication No. US20060089836A1
; GENERAL INFORMATION:
; APPLICANT: EXPRESSION DIAGNOSTICS, INC.
; APPLICANT: Wohlgenuth, Jay
; APPLICANT: Fry, Kirk
; APPLICANT: Woodward, Robert
; APPLICANT: Ly, Ngoc
; APPLICANT: Prentice, James
; APPLICANT: Morris, MacDonald
; APPLICANT: Rosenberg, Steven
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; TITLE OF INVENTION: AND MONITORING TRANSPLANT REJECTION
; FILE REFERENCE: 506612000104
; CURRENT APPLICATION NUMBER: US/10/511,937
; PRIOR FILING DATE: 2004-10-19
; PRIOR APPLICATION NUMBER: PCT/US2003/012946
; PRIOR FILING DATE: 2003-04-24
; PRIOR APPLICATION NUMBER: US 10/131,831
; PRIOR FILING DATE: 2002-04-24
; PRIOR APPLICATION NUMBER: US 10/325,899
; PRIOR FILING DATE: 2002-12-20
; NUMBER OF SEQ ID NOS: 3117
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 485
; LENGTH: 2165
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-511-937-485
Query Match 81.2%; Score 13.8; DB 6; Length 2165;
Best Local Similarity 88.2%; Pred. No. 39;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCCGAG 17
Db 262 GCCCAGCTTGGCCGG 246

RESULT 2
US-10-511-937-485/c
; Sequence 485, Application US/10511937
; Publication No. US20060089836A1
; GENERAL INFORMATION:
; APPLICANT: EXPRESSION DIAGNOSTICS, INC.
; APPLICANT: Wohlgenuth, Jay
; APPLICANT: Fry, Kirk
; APPLICANT: Woodward, Robert
; APPLICANT: Ly, Ngoc
; APPLICANT: Prentice, James
; APPLICANT: Morris, MacDonald
; APPLICANT: Rosenberg, Steven
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; TITLE OF INVENTION: AND MONITORING TRANSPLANT REJECTION
; FILE REFERENCE: 506612000104
; CURRENT APPLICATION NUMBER: US/10/511,937
; PRIOR FILING DATE: 2004-10-19
; PRIOR APPLICATION NUMBER: PCT/US2003/012946
; PRIOR FILING DATE: 2003-04-24
; PRIOR APPLICATION NUMBER: US 10/131,831
; PRIOR FILING DATE: 2002-04-24
; PRIOR APPLICATION NUMBER: US 10/325,899
; PRIOR FILING DATE: 2002-12-20
; NUMBER OF SEQ ID NOS: 3117
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 485
; LENGTH: 2165
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-511-937-485
Query Match 81.2%; Score 13.8; DB 6; Length 2165;
Best Local Similarity 88.2%; Pred. No. 39;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCCGAG 17
Db 262 GCCCAGCTTGGCCGG 246

US-09-484-331-21
; Sequence 21, Application US/09484331
; Publication No. US20060105318A1
; GENERAL INFORMATION:
; APPLICANT: HARRINGTON, JOHN J.
; APPLICANT: SHERF, BRUCE
; APPLICANT: RUNDLETT, STEPHEN
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; TITLE OF INVENTION: ENDOGENOUS GENES
; FILE REFERENCE: 0221-0003L
; CURRENT APPLICATION NUMBER: US/09/484,331
; CURRENT FILING DATE: 2000-01-18
; PRIOR APPLICATION NUMBER: 09/276,820
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 09/263,814
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/253,022
; PRIOR FILING DATE: 1999-02-19
; PRIOR APPLICATION NUMBER: 09/159,643
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 08/941,223
; PRIOR FILING DATE: 1997-09-26
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 5382
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (890)
; OTHER INFORMATION: a, c, t, g, other or unknown
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1042)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-09-484-331-21
Query Match 81.2%; Score 13.8; DB 1; Length 5382;
Best Local Similarity 88.2%; Pred. No. 41;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCCGAG 17
Db 1501 GCCCGCATGCGCGAG 1517

RESULT 3
US-10-485-397-7
; Sequence 7, Application US/10485397
; Publication No. US20060099673A1
; GENERAL INFORMATION:
; APPLICANT: ALTANA Pharma AG
; TITLE OF INVENTION: Novel Recombinant Gene Expression Method
; FILE REFERENCE: B697USPCT01
; CURRENT APPLICATION NUMBER: US/10/485,397
; CURRENT FILING DATE: 2004-01-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 7873
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: pGFPstopneo is a circular Plasmid DNA
US-10-485-397-7
Query Match 81.2%; Score 13.8; DB 6; Length 7873;
Best Local Similarity 88.2%; Pred. No. 41;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GCCCAGCTTGGCCGAG 17
```

```
Db      5129 GCCCGCATGCCGCGAG 5145

RESULT 4
US-10-485-397-8
; Sequence 8, Application US/10485397
; Publication No. US20060099673A1
; GENERAL INFORMATION:
; APPLICANT: ALTANA Pharma AG
; TITLE OF INVENTION: Novel Recombinant Gene Expression Method
; FILE REFERENCE: B697USCT01
; CURRENT APPLICATION NUMBER: US/10/485,397
; CURRENT FILING DATE: 2004-01-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 7943
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: pGFPstopneosecis is a circular Plasmid DNA
US-10-485-397-8

Query Match      81.2%; Score 13.8; DB 6; Length 7943;
Best Local Similarity 88.2%; Pred. No. 41;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 GCCCAGCTTGCCGCGAG 17
      ||||| ||| ||||| |||||
Db      5199 GCCCGCATGCCGCGAG 5215

RESULT 5
US-09-484-331-22
; Sequence 22, Application US/09484331
; Publication No. US20060105318A1
; GENERAL INFORMATION:
; APPLICANT: HARRINGTON, JOHN J.
; APPLICANT: SHERP, BRUCE
; APPLICANT: RUNDLETT, STEPHEN
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; FILE REFERENCE: 0221-0003L
; CURRENT APPLICATION NUMBER: US/09/484,331
; CURRENT FILING DATE: 2000-01-18
; PRIOR APPLICATION NUMBER: 09/276,820
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 09/263,814
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/253,022
; PRIOR FILING DATE: 1999-02-19
; PRIOR APPLICATION NUMBER: 09/159,643
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 08/941,223
; PRIOR FILING DATE: 1997-09-26
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 9737
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8347)
; OTHER INFORMATION: a, c, t, g, other or unknown
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8499)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-09-484-331-22

Query Match      81.2%; Score 13.8; DB 1; Length 9737;
Best Local Similarity 88.2%; Pred. No. 42;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 GCCCAGCTTGCCGCGAG 17
      ||||| ||| ||||| |||||
Db      8958 GCCCGCATGCCGCGAG 8974

RESULT 6
US-09-484-331-23
; Sequence 23, Application US/09484331
; Publication No. US20060105318A1
; GENERAL INFORMATION:
; APPLICANT: HARRINGTON, JOHN J.
; APPLICANT: SHERP, BRUCE
; APPLICANT: RUNDLETT, STEPHEN
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; FILE REFERENCE: 0221-0003L
; CURRENT APPLICATION NUMBER: US/09/484,331
; CURRENT FILING DATE: 2000-01-18
; PRIOR APPLICATION NUMBER: 09/276,820
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 09/263,814
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/253,022
; PRIOR FILING DATE: 1999-02-19
; PRIOR APPLICATION NUMBER: 09/159,643
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 08/941,223
; PRIOR FILING DATE: 1997-09-26
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 23
; LENGTH: 9737
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8347)
; OTHER INFORMATION: a, c, t, g, other or unknown
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8499)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-09-484-331-23

Query Match      81.2%; Score 13.8; DB 1; Length 9737;
Best Local Similarity 88.2%; Pred. No. 42;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 GCCCAGCTTGCCGCGAG 17
      ||||| ||| ||||| |||||
Db      8958 GCCCGCATGCCGCGAG 8974

RESULT 7
US-09-484-331-28
; Sequence 28, Application US/09484331
; Publication No. US20060105318A1
; GENERAL INFORMATION:
; APPLICANT: HARRINGTON, JOHN J.
; APPLICANT: SHERP, BRUCE
; APPLICANT: RUNDLETT, STEPHEN
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; FILE REFERENCE: 0221-0003L
; CURRENT APPLICATION NUMBER: US/09/484,331
; CURRENT FILING DATE: 2000-01-18
; PRIOR APPLICATION NUMBER: 09/276,820
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 09/263,814
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/253,022
```

```
; PRIOR FILING DATE: 1999-02-19
; PRIOR APPLICATION NUMBER: 09/159,643
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 08/941,223
; PRIOR FILING DATE: 1997-09-26
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
; LENGTH: 9737
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8347)
; OTHER INFORMATION: a, c, t, g, other or unknown
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8499)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-09-484-331-28

Query Match      81.2%; Score 13.8; DB 1; Length 9737;
Best Local Similarity 88.2%; Pred. No. 42;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1  GCCCAGCGTTGGCCGAG 17
Db      8958  GCCCGCGCATGGCCGAG 8974

RESULT 8
US-09-484-331-24
; Sequence 24, Application US/09484331
; Publication No. US20060105318A1
; GENERAL INFORMATION:
; APPLICANT: HARRINGTON, JOHN J.
; APPLICANT: SHERF, BRUCE
; APPLICANT: RUNDLETT, STEPHEN
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; TITLE OF INVENTION: ENDOGENOUS GENES
; FILE REFERENCE: 0221-0003L
; CURRENT APPLICATION NUMBER: US/09/484,331
; CURRENT FILING DATE: 2000-01-18
; PRIOR APPLICATION NUMBER: 09/276,820
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 09/263,814
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/253,022
; PRIOR FILING DATE: 1999-02-19
; PRIOR APPLICATION NUMBER: 09/159,643
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 08/941,223
; PRIOR FILING DATE: 1997-09-26
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24
; LENGTH: 9871
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8481)
; OTHER INFORMATION: a, c, t, g, other or unknown
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8633)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-09-484-331-24

Query Match      81.2%; Score 13.8; DB 1; Length 9871;
Best Local Similarity 88.2%; Pred. No. 42;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy      1  GCCCAGCGTTGGCCGAG 17
Db      9092  GCCCGCGCATGGCCGAG 9108

RESULT 9
US-09-484-331-25
; Sequence 25, Application US/09484331
; Publication No. US20060105318A1
; GENERAL INFORMATION:
; APPLICANT: HARRINGTON, JOHN J.
; APPLICANT: SHERF, BRUCE
; APPLICANT: RUNDLETT, STEPHEN
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; TITLE OF INVENTION: ENDOGENOUS GENES
; FILE REFERENCE: 0221-0003L
; CURRENT APPLICATION NUMBER: US/09/484,331
; CURRENT FILING DATE: 2000-01-18
; PRIOR APPLICATION NUMBER: 09/276,820
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 09/263,814
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/253,022
; PRIOR FILING DATE: 1999-02-19
; PRIOR APPLICATION NUMBER: 09/159,643
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 08/941,223
; PRIOR FILING DATE: 1997-09-26
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 25
; LENGTH: 10060
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8670)
; OTHER INFORMATION: a, c, t, g, other or unknown
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8822)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-09-484-331-25

Query Match      81.2%; Score 13.8; DB 1; Length 10060;
Best Local Similarity 88.2%; Pred. No. 42;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1  GCCCAGCGTTGGCCGAG 17
Db      9281  GCCCGCGCATGGCCGAG 9297

RESULT 10
US-10-953-349-15149/c
; Sequence 15149, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nickolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; TITLE OF INVENTION: ENCODED THERBY
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 15149
; LENGTH: 806
; TYPE: DNA
; ORGANISM: Glycine max
US-10-953-349-15149

Query Match      78.8%; Score 13.4; DB 6; Length 806;
```

Best Local Similarity 93.3%; Pred. No. 61;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 CCACGCTTGGCCGAG 17  
||| |||||  
Db 492 CCATGCTTGGCCGAG 478

RESULT 11  
US-10-953-349-35663/c  
; Sequence 35663, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 35663  
; LENGTH: 2019  
; TYPE: DNA  
; ORGANISM: Zea mays subsp. mays  
US-10-953-349-35663

Query Match 78.8%; Score 13.4; DB 6; Length 2019;  
Best Local Similarity 93.3%; Pred. No. 64;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 CCACGCTTGGCCGAG 17  
||| |||||  
Db 203 CCACGCTTGGCCGAG 189

RESULT 12  
US-11-293-697-559  
; Sequence 559, Application US/11293697  
; Publication No. US20060105376A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: Novel full length cDNA  
; FILE REFERENCE: H1-A0106  
; CURRENT APPLICATION NUMBER: US/11/293,697  
; CURRENT FILING DATE: 2005-12-05  
; PRIOR APPLICATION NUMBER: US/10/108,260  
; PRIOR FILING DATE: 2002-03-28  
; NUMBER OF SEQ ID NOS: 5458  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 559  
; LENGTH: 2175  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-293-697-559

Query Match 78.8%; Score 13.4; DB 7; Length 2175;  
Best Local Similarity 93.3%; Pred. No. 64;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 CCACGCTTGGCCGAG 17  
||| |||||  
Db 1948 CCACGCTTGGCCGAG 1962

RESULT 13  
US-10-511-937-648/c  
; Sequence 648, Application US/10511937  
; Publication No. US20060088836A1  
; GENERAL INFORMATION:  
; APPLICANT: EXPRESSION DIAGNOSTICS, INC.  
; APPLICANT: Wohlgenuth, Jay  
; APPLICANT: Fry, Kirk

; APPLICANT: Woodward, Robert  
; APPLICANT: Ly, Ngoc  
; APPLICANT: Prentice, James  
; APPLICANT: Morris, Macdonald  
; APPLICANT: Rosenberg, Steven  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING  
; TITLE OF INVENTION: AND MONITORING TRANSPLANT REJECTION  
; FILE REFERENCE: 506612000104  
; CURRENT APPLICATION NUMBER: US/10/511,937  
; CURRENT FILING DATE: 2004-10-19  
; PRIOR APPLICATION NUMBER: PCT/US2003/012946  
; PRIOR FILING DATE: 2003-04-24  
; PRIOR APPLICATION NUMBER: US 10/131,831  
; PRIOR FILING DATE: 2002-04-24  
; PRIOR APPLICATION NUMBER: US 10/325,899  
; PRIOR FILING DATE: 2002-12-20  
; NUMBER OF SEQ ID NOS: 3117  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 648  
; LENGTH: 2811  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-511-937-648

Query Match 78.8%; Score 13.4; DB 6; Length 2811;  
Best Local Similarity 93.3%; Pred. No. 65;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 CCACGCTTGGCCGAG 17  
||| |||||  
Db 1465 CCACCTTGTGGCCGAG 1451

RESULT 14  
US-11-293-697-950  
; Sequence 950, Application US/11293697  
; Publication No. US20060105376A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: Novel full length cDNA  
; FILE REFERENCE: H1-A0106  
; CURRENT APPLICATION NUMBER: US/11/293,697  
; CURRENT FILING DATE: 2005-12-05  
; PRIOR APPLICATION NUMBER: US/10/108,260  
; PRIOR FILING DATE: 2002-03-28  
; NUMBER OF SEQ ID NOS: 5458  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 950  
; LENGTH: 3048  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-293-697-950

Query Match 78.8%; Score 13.4; DB 7; Length 3048;  
Best Local Similarity 93.3%; Pred. No. 65;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCCACGCTTGGCCGA 16  
||| |||||  
Db 838 CCCACGCTTGGCCAA 852

RESULT 15  
US-11-169-140-80  
; Sequence 80, Application US/11169140  
; Publication No. US20060099150A1  
; GENERAL INFORMATION:  
; APPLICANT: ARIZKE PHARMACEUTICALS, INC.  
; APPLICANT: HOUSTON, Lou, L.  
; APPLICANT: SHAWLEY, Philip, J.  
; APPLICANT: HAWLEY, Stephen  
; APPLICANT: GLYNN, Jacqueline, M.  
; APPLICANT: CHAPIN, Steven



```
RESULT 19
US-11-249-305-2
; Sequence 2, Application US/11249305
; Publication No. US20060115836A1
; GENERAL INFORMATION:
; APPLICANT: TRINQUET, VALERIE
; APPLICANT: ROSSIO, PATRICIA
; APPLICANT: FOGEL, PAUL
; TITLE OF INVENTION: UTILIZATION OF PANCC, DAD1, GRIM19 AND HADH1 GENES FOR
; FILE REFERENCE: 034227-098
; CURRENT APPLICATION NUMBER: US/11/249,305
; PRIOR FILING DATE: 2005-10-14
; PRIOR APPLICATION NUMBER: PCT/FR04/000934
; PRIOR FILING DATE: 2004-04-15
; PRIOR APPLICATION NUMBER: FR 0304773
; PRIOR FILING DATE: 2003-04-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 2
; LENGTH: 500
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-249-305-2

Query Match          75.3%; Score 12.8; DB 7; Length 500;
Best Local Similarity 87.5%; Pred. No. 1.3e+02;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GCCCAGCGCTTGGCCGA 16
Db 424 GCCCAGCGCTTGGCCGA 439

RESULT 20
US-11-249-305-3/c
; Sequence 3, Application US/11249305
; Publication No. US20060115836A1
; GENERAL INFORMATION:
; APPLICANT: TRINQUET, VALERIE
; APPLICANT: ROSSIO, PATRICIA
; APPLICANT: FOGEL, PAUL
; TITLE OF INVENTION: UTILIZATION OF PANCC, DAD1, GRIM19 AND HADH1 GENES FOR
; FILE REFERENCE: 034227-098
; CURRENT APPLICATION NUMBER: US/11/249,305
; PRIOR FILING DATE: 2005-10-14
; PRIOR APPLICATION NUMBER: PCT/FR04/000934
; PRIOR FILING DATE: 2004-04-15
; PRIOR APPLICATION NUMBER: FR 0304773
; PRIOR FILING DATE: 2003-04-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 3
; LENGTH: 500
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-249-305-3

Query Match          75.3%; Score 12.8; DB 7; Length 500;
Best Local Similarity 87.5%; Pred. No. 1.3e+02;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GCCCAGCGCTTGGCCGA 16
Db 77 GCCCAGCGCTTGGCCGA 62

RESULT 21
US-11-249-305-4
; Sequence 4, Application US/11249305
; Publication No. US20060115836A1
```

```
; GENERAL INFORMATION:
; APPLICANT: TRINQUET, VALERIE
; APPLICANT: ROSSIO, PATRICIA
; APPLICANT: FOGEL, PAUL
; TITLE OF INVENTION: UTILIZATION OF PANCC, DAD1, GRIM19 AND HADH1 GENES FOR
; FILE REFERENCE: 034227-098
; CURRENT APPLICATION NUMBER: US/11/249,305
; PRIOR FILING DATE: 2005-10-14
; PRIOR APPLICATION NUMBER: PCT/FR04/000934
; PRIOR FILING DATE: 2004-04-15
; PRIOR APPLICATION NUMBER: FR 0304773
; PRIOR FILING DATE: 2003-04-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 4
; LENGTH: 500
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-249-305-4

Query Match          75.3%; Score 12.8; DB 7; Length 500;
Best Local Similarity 75.0%; Pred. No. 1.3e+02;
Matches 12; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GCCCAGCGCTTGGCCGA 16
Db 424 GCCCAGCGCTTGGCCGA 439

RESULT 22
US-11-217-529-82759
; Sequence 82759, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 82759
; LENGTH: 693
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-82759

Query Match          75.3%; Score 12.8; DB 7; Length 693;
Best Local Similarity 87.5%; Pred. No. 1.3e+02;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 CCCACGCTTGGCCGAG 17
Db 534 CCCACGCTTGGCCGAG 549

RESULT 23
US-10-953-349-20941/c
; Sequence 20941, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nickolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1579PUS2
```

```
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 20941
; LENGTH: 726
; TYPE: DNA
; ORGANISM: Glycine max
US-10-953-349-20941

Query Match          75.3%; Score 12.8; DB 6; Length 726;
Best Local Similarity 87.5%; Pred. No. 1.3e+02;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 CCCACGCTTGCCCGAG 17
Db 395 CCCGCGCTTGGCGCAG 380

RESULT 24
US-11-217-529-82116/c
; Sequence 82116, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 82116
; LENGTH: 777
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-82116

Query Match          75.3%; Score 12.8; DB 7; Length 777;
Best Local Similarity 87.5%; Pred. No. 1.3e+02;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 CCCACGCTTGCCCGAG 17
Db 125 CCCACACTTGCCCAAG 110

RESULT 25
US-11-217-529-1630
; Sequence 1630, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1630
```

```
; LENGTH: 1086
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-1630

Query Match          75.3%; Score 12.8; DB 7; Length 1086;
Best Local Similarity 87.5%; Pred. No. 1.3e+02;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GCCCAGCGTTGGCCGA 16
Db 24 GCCCAGCGTGGGCAGA 39

RESULT 26
US-10-953-349-25892/c
; Sequence 25892, Application US/109533349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 25892
; LENGTH: 1093
; TYPE: DNA
; ORGANISM: Triticum aestivum
US-10-953-349-25892

Query Match          75.3%; Score 12.8; DB 6; Length 1093;
Best Local Similarity 87.5%; Pred. No. 1.3e+02;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 CCCACGCTTGCCCGAG 17
Db 805 CCCACGTTGCCCGAG 790

RESULT 27
US-10-953-349-33000/c
; Sequence 33000, Application US/109533349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 33000
; LENGTH: 1124
; TYPE: DNA
; ORGANISM: Zea mays subsp. mays
US-10-953-349-33000

Query Match          75.3%; Score 12.8; DB 6; Length 1124;
Best Local Similarity 87.5%; Pred. No. 1.3e+02;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 CCCACGCTTGCCCGAG 17
Db 58 CCCATGCGTGCCCGAG 43

RESULT 28
US-10-953-349-35987
; Sequence 35987, Application US/109533349
```



```
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; TITLE OF INVENTION: ENCODED THERBY
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 35987
; LENGTH: 1541
; TYPE: DNA
; ORGANISM: Zea mays subsp. mays
US-10-953-349-35987

Query Match          75.3%; Score 12.8; DB 6; Length 1541;
Best Local Similarity 87.5%; Pred. No. 1.3e+02;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCCCAGCGTTGGCCGA 16
    ||||| ||||| |||||
Db 1372 GCCCAATCTGGCCGA 1387

RESULT 29
US-10-953-349-33688/c
; Sequence 33688, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; TITLE OF INVENTION: ENCODED THERBY
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 33688
; LENGTH: 1824
; TYPE: DNA
; ORGANISM: Zea mays subsp. mays
US-10-953-349-33688

Query Match          75.3%; Score 12.8; DB 6; Length 1824;
Best Local Similarity 87.5%; Pred. No. 1.3e+02;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CCCACGCTTGCCCGAG 17
    ||||| ||||| |||||
Db 1088 CCCATGCTAGGCCGAG 1073

RESULT 30
US-11-293-697-1055
; Sequence 1055, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1055
; LENGTH: 2525
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-293-697-1055
```

```
Query Match          75.3%; Score 12.8; DB 7; Length 2525;
Best Local Similarity 87.5%; Pred. No. 1.4e+02;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CCCACGCTTGCCCGAG 17
    ||||| ||||| |||||
Db 110 CCCACGCTTGCCCGAG 125

RESULT 31
US-11-293-697-1132/c
; Sequence 1132, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1132
; LENGTH: 2557
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-293-697-1132

Query Match          75.3%; Score 12.8; DB 7; Length 2557;
Best Local Similarity 87.5%; Pred. No. 1.4e+02;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CCCACGCTTGCCCGAG 17
    ||||| ||||| |||||
Db 876 CCCACGCTTGCCCGAG 861

RESULT 32
US-11-293-697-845
; Sequence 845, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 845
; LENGTH: 2593
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-293-697-845

Query Match          75.3%; Score 12.8; DB 7; Length 2593;
Best Local Similarity 87.5%; Pred. No. 1.4e+02;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CCCACGCTTGCCCGAG 17
    ||||| ||||| |||||
Db 1291 CCCTGCTTGCCCGAG 1306

RESULT 33
US-11-293-697-1461/c
; Sequence 1461, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
```

```
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: HI-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1461
; LENGTH: 2641
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-293-697-1461

Query Match          75.3%; Score 12.8; DB 7; Length 2641;
Best Local Similarity 87.5%; Pred. No. 1.4e+02;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 CCCACGCTTGCCCGAG 17
    ||||| ||||| |||||
Db 118 CCCAAGCTTGGCTGAG 133

RESULT 36
US-11-293-697-291
; Sequence 291, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: HI-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 291
; LENGTH: 3095
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-293-697-291

Query Match          75.3%; Score 12.8; DB 7; Length 3095;
Best Local Similarity 87.5%; Pred. No. 1.4e+02;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 CCCACGCTTGCCCGAG 17
    ||||| ||||| |||||
Db 2952 CCCACCTGGGCCGAG 2967

RESULT 37
US-11-254-185-1
; Sequence 1, Application US/11254185
; Publication No. US20060099625A1
; GENERAL INFORMATION:
; APPLICANT: CRAIK, CHARLES S.
; APPLICANT: TAKEUCHI, TOSHIHIKO
; APPLICANT: SCHUMAN, MARC
; TITLE OF INVENTION: MT-SPI SERINE PROTEASE
; FILE REFERENCE: 28644-701.302
; CURRENT APPLICATION NUMBER: US/11/254,185
; CURRENT FILING DATE: 2005-10-18
; NUMBER OF SEQ ID NOS: 83
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1
; LENGTH: 3142
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (37)..(2601)
US-11-254-185-1

Query Match          75.3%; Score 12.8; DB 7; Length 3142;
Best Local Similarity 87.5%; Pred. No. 1.4e+02;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 CCCACGCTTGCCCGAG 17
    ||||| ||||| |||||
Db 2876 CCCAAGCTGGGCCGAG 2891

RESULT 38
US-11-253-869-1
; Sequence 1, Application US/11253869
; Publication No. US20060104979A1
```

```

; GENERAL INFORMATION:
; APPLICANT: CRAIK, CHARLES S.
; APPLICANT: TAKEUCHI, TOSHIHIKO
; APPLICANT: SCHUMAN, MARC
; TITLE OF INVENTION: MT-SPI POLYNUCLEOTIDES AND POLYPEPTIDES
; FILE REFERENCE: 28644-701.303
; CURRENT APPLICATION NUMBER: US/11/253,869
; CURRENT FILING DATE: 2005-10-18
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1
; LENGTH: 3142
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (37)..(2601)
US-11-253-869-1

Query Match          75.3%; Score 12.8; DB 7; Length 3142;
Best Local Similarity 87.5%; Pred. No. 1.4e+02;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CCCACGCTTGCCGAG 17
Db 2876 CCCACGCTTGCCGAG 2891
||||| ||| |||||

RESULT 39
US-11-293-697-510
; Sequence 510, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 510
; LENGTH: 3693
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-293-697-510

Query Match          75.3%; Score 12.8; DB 7; Length 3693;
Best Local Similarity 87.5%; Pred. No. 1.4e+02;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CCCACGCTTGCCGAG 17
Db 1760 CCCACGCTTGCCGAG 1775
||||| ||| |||||

RESULT 40
US-11-121-154-158
; Sequence 158, Application US/1121154
; Publication No. US20060105914A1
; GENERAL INFORMATION:
; APPLICANT: TAYLOR, LARRY EDMUND
; APPLICANT: WEINER, RONALD M.
; APPLICANT: HUTCHESON, STEVEN WAYNE
; APPLICANT: EKBORG, NATHAN A.
; APPLICANT: HOWARD, MICHAEL
; TITLE OF INVENTION: PLANT WALL DEGRADATIVE COMPOUNDS AND SYSTEMS
; FILE REFERENCE: 108172-00121
; CURRENT APPLICATION NUMBER: US/11/121,154
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,971
; PRIOR FILING DATE: 2004-05-04
```

```

; NUMBER OF SEQ ID NOS: 214
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 158
; LENGTH: 5202
; TYPE: DNA
; ORGANISM: Microbulbifer degradans
US-11-121-154-158

Query Match          75.3%; Score 12.8; DB 7; Length 5202;
Best Local Similarity 87.5%; Pred. No. 1.4e+02;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCCCAGCTTGCCGCA 16
Db 4889 GCCCAGCTTGCCGCA 4904
||||| ||| |||||

RESULT 41
US-11-217-529-1662
; Sequence 1662, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1662
; LENGTH: 9426
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-1662

Query Match          75.3%; Score 12.8; DB 7; Length 9426;
Best Local Similarity 87.5%; Pred. No. 1.4e+02;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCCCAGCTTGCCGCA 16
Db 8285 GCCCAGCTTGCTGA 8300
||||| ||| |||||

RESULT 42
US-10-505-928-568/c
; Sequence 568, Application US/10505928
; Publication No. US20060088532A1
; GENERAL INFORMATION:
; APPLICANT: Ludwig Institute for Cancer Research et al.
; TITLE OF INVENTION: LYMPHATIC ENDOTHELIAL GENES
; FILE REFERENCE: 28967/39178
; CURRENT APPLICATION NUMBER: US/10/505,928
; CURRENT FILING DATE: 2004-08-27
; PRIOR APPLICATION NUMBER: US 60/363,019
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 866
; SOFTWARE: PatentIn 3.2
; SEQ ID NO 568
; LENGTH: 10905
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-505-928-568

Query Match          75.3%; Score 12.8; DB 6; Length 10905;
Best Local Similarity 87.5%; Pred. No. 1.5e+02;
```

Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 CCCACGCTTGCCCGAG 17  
Db 774 CCCACGCTTGCCCGAG 759

RESULT 43

US-10-857-260-34  
; Sequence 34, Application US/10857260  
; Publication No. US20060110742A1  
; GENERAL INFORMATION:  
; APPLICANT: Lyons, Leslie A.  
; APPLICANT: Grahn, Robert  
; APPLICANT: Erdman, Carolyn  
; APPLICANT: The Regents of the University of California  
; TITLE OF INVENTION: Carrier Tests for Polycystic Kidney Disease in the Cat  
; FILE REFERENCE: 023070-146800US  
; CURRENT APPLICATION NUMBER: US/10/857,260  
; CURRENT FILING DATE: 2004-05-28  
; NUMBER OF SEQ ID NOS: 47  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 34  
; LENGTH: 27419  
; TYPE: DNA  
; ORGANISM: Felis catus  
; FEATURE:  
; OTHER INFORMATION: GenBank Accession No. AC145332.28 genomic DNA for  
; OTHER INFORMATION: PKD1 27419 bp contig  
; FEATURE:  
; NAME/KEY: modified\_base  
; LOCATION: (1)..(27419)  
; OTHER INFORMATION: n = g, a, c or t  
US-10-857-260-34

Query Match 75.3%; Score 12.8; DB 6; Length 27419;  
Best Local Similarity 87.5%; Pred. No. 1.5e+02;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 CCCACGCTTGCCCGAG 17  
Db 16104 CCCCCTTGCCCGAG 16119

RESULT 44

US-10-505-928-151  
; Sequence 151, Application US/10505928  
; Publication No. US20060088532A1  
; GENERAL INFORMATION:  
; APPLICANT: Ludwig Institute for Cancer Research et al.  
; TITLE OF INVENTION: LYMPHATIC ENDOTHELIAL GENES  
; FILE REFERENCE: 28967/39178  
; CURRENT APPLICATION NUMBER: US/10/505,928  
; CURRENT FILING DATE: 2004-08-27  
; PRIOR FILING DATE: 2002-03-07  
; NUMBER OF SEQ ID NOS: 866  
; SOFTWARE: PatentIn 3.2  
; SEQ ID NO 151  
; LENGTH: 128361  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-505-928-151

Query Match 75.3%; Score 12.8; DB 6; Length 128361;  
Best Local Similarity 87.5%; Pred. No. 1.5e+02;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 CCCACGCTTGCCCGAG 17  
Db 46066 CCCACACTTGCCCAAG 46081

RESULT 45

US-10-489-730-10  
; GENERAL INFORMATION:  
; APPLICANT: Melino, Gennaro  
; APPLICANT: Hayes, Ian  
; APPLICANT: de Laurenzi, Vincenzo  
; APPLICANT: Barcarolli, Daniela  
; APPLICANT: Candi, Eleonora  
; APPLICANT: Bernassola, Francesca  
; APPLICANT: Tobler, Andreas  
; APPLICANT: Novak, Urban  
; TITLE OF INVENTION: Human Delta-N p73 Molecules and Uses Thereof  
; FILE REFERENCE: 19319.002  
; CURRENT APPLICATION NUMBER: US/10/489,730  
; CURRENT FILING DATE: 2004-03-16  
; PRIOR APPLICATION NUMBER: PCT/GB02/04238  
; PRIOR FILING DATE: 2002-09-17  
; PRIOR APPLICATION NUMBER: US 60/322,436  
; PRIOR FILING DATE: 2001-09-17  
; NUMBER OF SEQ ID NOS: 39  
; OTHER INFORMATION: reverse complement of exons 14 through 1 as reported in GenBank A  
US-10-489-730-10

Query Match 75.3%; Score 12.8; DB 6; Length 138941;  
Best Local Similarity 87.5%; Pred. No. 1.5e+02;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GCCCACGCTTGCCCGA 16  
Db 116980 GCCCACGAGGCCGA 116995

RESULT 46

US-11-169-140-82  
; Sequence 82, Application US/11169140  
; Publication No. US20060099150A1  
; GENERAL INFORMATION:  
; APPLICANT: ARIZEKE PHARMACEUTICALS, INC.  
; APPLICANT: HOUSTON, Lou, L.  
; APPLICANT: SHERIDAN, Philip, J.  
; APPLICANT: HAWLEY, Stephen  
; APPLICANT: GLYNN, Jacqueline, M.  
; APPLICANT: CHAPIN, Steven  
; TITLE OF INVENTION: METHODS FOR THE TRANSPORT OF BIOLOGICALLY ACTIVE AGENTS ACROSS CE  
; FILE REFERENCE: 11474-037-999  
; CURRENT APPLICATION NUMBER: US/11/169,140  
; CURRENT FILING DATE: 2005-06-27  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: US 60/267,601  
; PRIOR FILING DATE: 2000-11-14  
; PRIOR APPLICATION NUMBER: US 60/248,478  
; PRIOR FILING DATE: 2000-11-13  
; PRIOR APPLICATION NUMBER: US 60/237,929  
; PRIOR FILING DATE: 2000-10-02  
; PRIOR APPLICATION NUMBER: 09/949,039  
; PRIOR FILING DATE: 2001-09-06  
; PRIOR APPLICATION NUMBER: 09/969,748  
; PRIOR FILING DATE: 2001-10-02  
; NUMBER OF SEQ ID NOS: 143  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 82  
; LENGTH: 39  
; TYPE: DNA  
; ORGANISM: Murinae gen. sp.  
US-11-169-140-82

Query Match 72.9%; Score 12.4; DB 7; Length 39;  
Best Local Similarity 92.9%; Pred. No. 1.9e+02;  
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCACGCTTGCC 14

Db 1 GCCCAAGCTTGCC 14  
||||| |||||

## RESULT 47

US-11-169-140-87  
; Sequence 87, Application US/11169140  
; Publication No. US20060099150A1  
; GENERAL INFORMATION:  
; APPLICANT: ARIZEKE PHARMACEUTICALS, INC.  
; APPLICANT: HOUSTON, Lou, L.  
; APPLICANT: SHERIDAN, Philip, J.  
; APPLICANT: HAWLEY, Stephen  
; APPLICANT: GLYNN, Jacqueline, M.  
; APPLICANT: CHAPIN, Steven

; TITLE OF INVENTION: METHODS FOR THE TRANSPORT OF BIOLOGICALLY ACTIVE AGENTS ACROSS CELL MEMBRANES  
; FILE REFERENCE: 11474-037-999  
; CURRENT FILING DATE: 2005-06-27  
; PRIOR FILING DATE: 2005-06-27  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: US 60/267,601  
; PRIOR APPLICATION NUMBER: US 60/248,819  
; PRIOR FILING DATE: 2000-11-14  
; PRIOR APPLICATION NUMBER: US 60/248,478  
; PRIOR FILING DATE: 2000-11-13  
; PRIOR APPLICATION NUMBER: US 60/237,929  
; PRIOR FILING DATE: 2000-10-02  
; PRIOR APPLICATION NUMBER: 09/949,039  
; PRIOR FILING DATE: 2001-09-06  
; PRIOR APPLICATION NUMBER: 09/969,748  
; PRIOR FILING DATE: 2001-10-02  
; NUMBER OF SEQ ID NOS: 143  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 87  
; LENGTH: 39  
; TYPE: DNA  
; ORGANISM: Murinae gen. sp.

US-11-169-140-87

Query Match 72.9%; Score 12.4; DB 7; Length 39;  
Best Local Similarity 92.9%; Pred. No. 1.9e+02;  
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCCAGCTTGCC 14  
||||| |||||

## RESULT 48

US-11-169-140-95  
; Sequence 95, Application US/11169140  
; Publication No. US20060099150A1  
; GENERAL INFORMATION:  
; APPLICANT: ARIZEKE PHARMACEUTICALS, INC.  
; APPLICANT: HOUSTON, Lou, L.  
; APPLICANT: SHERIDAN, Philip, J.  
; APPLICANT: HAWLEY, Stephen  
; APPLICANT: GLYNN, Jacqueline, M.  
; APPLICANT: CHAPIN, Steven

; TITLE OF INVENTION: METHODS FOR THE TRANSPORT OF BIOLOGICALLY ACTIVE AGENTS ACROSS CELL MEMBRANES  
; FILE REFERENCE: 11474-037-999  
; CURRENT APPLICATION NUMBER: US/11/169,140  
; CURRENT FILING DATE: 2005-06-27  
; PRIOR FILING DATE: 2005-06-27  
; PRIOR APPLICATION NUMBER: US 60/267,601  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: US 60/248,819  
; PRIOR FILING DATE: 2000-11-14  
; PRIOR APPLICATION NUMBER: US 60/248,478  
; PRIOR FILING DATE: 2000-11-13  
; PRIOR APPLICATION NUMBER: US 60/237,929  
; PRIOR FILING DATE: 2000-10-02

; PRIOR APPLICATION NUMBER: 09/949,039  
; PRIOR FILING DATE: 2001-09-06  
; PRIOR APPLICATION NUMBER: 09/969,748  
; PRIOR FILING DATE: 2001-10-02  
; NUMBER OF SEQ ID NOS: 143  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 95  
; LENGTH: 39  
; TYPE: DNA  
; ORGANISM: Rattus sp.  
; US-11-169-140-95

Query Match 72.9%; Score 12.4; DB 7; Length 39;  
Best Local Similarity 92.9%; Pred. No. 1.9e+02;  
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCCAGCTTGCC 14  
||||| |||||

## RESULT 49

US-10-953-349-26326  
; Sequence 26326, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nikolai et al.

; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 26326  
; LENGTH: 506  
; TYPE: DNA  
; ORGANISM: Triticum aestivum  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (477)..(477)  
; OTHER INFORMATION: n is a, c, g, or t  
US-10-953-349-26326

Query Match 72.9%; Score 12.4; DB 6; Length 506;  
Best Local Similarity 92.9%; Pred. No. 2.1e+02;  
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCCACGCTTGCCG 15  
||||| |||||

## RESULT 50

US-10-953-349-37335/c  
; Sequence 37335, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nikolai et al.

; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 37335  
; LENGTH: 1035  
; TYPE: DNA  
; ORGANISM: Zea mays subsp. mays  
US-10-953-349-37335

Query Match 72.9%; Score 12.4; DB 6; Length 1035;

Best Local Similarity 92.9%; Pred. No. 2.2e+02; Indels 0; Gaps 0;  
Matches 13; Conservative 0; Mismatches 1;

QY 1 GCCACGCTTGCC 14  
Db 324 GCCACGCTTGCC 311

## RESULT 51

US-11-217-529-272/c  
; Sequence 272, Application US/11217529  
; Publication No. US20060099612A1  
; GENERAL INFORMATION:  
; APPLICANT: SUNTORY LIMITED  
; APPLICANT: NAKAO, YOSHIHIRO  
; APPLICANT: NAKAMURA, NORIHO  
; APPLICANT: KODAMA, YUKIO  
; APPLICANT: FUJIMURA, TOMOKO  
; APPLICANT: ASHIKARI, TOSHIHIKO  
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS  
; FILE REFERENCE: S-38-285  
; CURRENT APPLICATION NUMBER: US 10/932,182  
; CURRENT FILING DATE: 2005-09-02  
; PRIOR APPLICATION NUMBER: US 10/932,182  
; PRIOR FILING DATE: 2004-09-02  
; NUMBER OF SEQ ID NOS: 197023  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 272  
; LENGTH: 1242  
; TYPE: DNA  
; ORGANISM: Saccharomyces pastorianus

## US-11-217-529-272

Query Match 72.9%; Score 12.4; DB 7; Length 1242;  
Best Local Similarity 92.9%; Pred. No. 2.2e+02;  
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCACGCTTGCC 14  
Db 120 GCCACGCTTGCC 107

## RESULT 52

US-10-471-571A-2011  
; Sequence 2011, Application US/10471571A  
; Publication No. US20060115490A1  
; GENERAL INFORMATION:  
; APPLICANT: CHIRON SPA  
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS PROTEINS AND NUCLEIC ACIDS  
; FILE REFERENCE: P026927WO  
; CURRENT APPLICATION NUMBER: US/10/471,571A  
; CURRENT FILING DATE: 2003-09-12  
; PRIOR APPLICATION NUMBER: GB-0107661.1  
; PRIOR FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 5642  
; SOFTWARE: SeqWin99, version 1.03  
; SEQ ID NO 2011  
; LENGTH: 1305  
; TYPE: DNA  
; ORGANISM: Staphylococcus aureus

## US-10-471-571A-2011

Query Match 72.9%; Score 12.4; DB 6; Length 1305;  
Best Local Similarity 92.9%; Pred. No. 2.2e+02;  
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCACGCTTGCC 14  
Db 75 GCCACGCTTGCC 88

## RESULT 53

US-10-953-349-35503/c

; Sequence 35503, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nickolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 35503  
; LENGTH: 1423  
; TYPE: DNA  
; ORGANISM: Zea mays subsp. mays  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (513)..(513)  
; OTHER INFORMATION: n is a, c, g, or t  
US-10-953-349-35503

Query Match 72.9%; Score 12.4; DB 6; Length 1423;  
Best Local Similarity 92.9%; Pred. No. 2.2e+02;  
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 CCACGCTTGCCGA 16  
Db 81 CCACGCTAGGCCGA 68

## RESULT 54

US-10-953-349-35104  
; Sequence 35104, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nickolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 35104  
; LENGTH: 1429  
; TYPE: DNA  
; ORGANISM: Zea mays subsp. mays  
US-10-953-349-35104

Query Match 72.9%; Score 12.4; DB 6; Length 1429;  
Best Local Similarity 92.9%; Pred. No. 2.2e+02;  
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 CACGCTTGCCGAG 17  
Db 1189 CACGATGCCGAG 1202

## RESULT 55

US-11-293-697-3/c  
; Sequence 3, Application US/11293697  
; Publication No. US20060105376A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: Novel full length cDNA  
; FILE REFERENCE: H1-A0106  
; CURRENT APPLICATION NUMBER: US/11/293,697  
; CURRENT FILING DATE: 2005-12-05  
; PRIOR APPLICATION NUMBER: US/10/108,260  
; PRIOR FILING DATE: 2002-03-28  
; NUMBER OF SEQ ID NOS: 5458  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3

```
; LENGTH: 1685
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-293-697-3

Query Match      72.9%; Score 12.4; DB 7; Length 1685;
Best Local Similarity 92.9%; Pred. No. 2.2e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      4 CACGCTTGCCGAG 17
Db      601 CACCTTGCCGAG 588

RESULT 56
US-11-293-697-559/c
; Sequence 559, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 559
; LENGTH: 2175
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-293-697-559

Query Match      72.9%; Score 12.4; DB 7; Length 2175;
Best Local Similarity 92.9%; Pred. No. 2.2e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 GCCCAGCGCTTGCC 14
Db      817 GCCCAGCGCTTGCC 804

RESULT 57
US-11-293-697-322
; Sequence 322, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 322
; LENGTH: 2299
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-293-697-322

Query Match      72.9%; Score 12.4; DB 7; Length 2299;
Best Local Similarity 92.9%; Pred. No. 2.2e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      4 CACGCTTGCCGAG 17
Db      417 CACGCTTGCCGAG 430

RESULT 58
US-11-293-697-1089
; Sequence 1089, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1089
; LENGTH: 2728
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-293-697-1089

Query Match      72.9%; Score 12.4; DB 7; Length 2728;
Best Local Similarity 92.9%; Pred. No. 2.3e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 GCCCAGCGCTTGCC 14
Db      573 GCCCAGCGCTTGCC 586

RESULT 59
US-11-293-697-1682/c
; Sequence 1682, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1682
; LENGTH: 2750
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-293-697-1682

Query Match      72.9%; Score 12.4; DB 7; Length 2750;
Best Local Similarity 92.9%; Pred. No. 2.3e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 GCCCAGCGCTTGCC 14
Db      187 GCCCAGCGCTTGCC 174

RESULT 60
US-10-511-937-576/c
; Sequence 576, Application US/10511937
; Publication No. US2006008836A1
; GENERAL INFORMATION:
; APPLICANT: EXPRESSION DIAGNOSTICS, INC.
; APPLICANT: Wohlgenuth, Jay
; APPLICANT: Fry, Kirk
; APPLICANT: Woodward, Robert
; APPLICANT: Ly, Ngoc
; APPLICANT: Prentice, James
; APPLICANT: Morris, MacDonald
; APPLICANT: Rosenberg, Steven
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; FILE REFERENCE: AND MONITORING TRANSPLANT REJECTION
; FILE REFERENCE: 506612000104
```

```
/ CURRENT APPLICATION NUMBER: US/10/511,937
/ CURRENT FILING DATE: 2004-10-19
/ PRIOR APPLICATION NUMBER: PCT/US2003/012946
/ PRIOR FILING DATE: 2003-04-24
/ PRIOR APPLICATION NUMBER: US 10/131,831
/ PRIOR FILING DATE: 2002-04-24
/ PRIOR APPLICATION NUMBER: US 10/325,899
/ PRIOR FILING DATE: 2002-12-20
/ NUMBER OF SEQ ID NOS: 3117
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 576
/ LENGTH: 2882
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-511-937-576

Query Match          72.9%; Score 12.4; DB 6; Length 2882;
Best Local Similarity 92.9%; Pred. No. 2.3e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2  CCACGCTTGGCCG 15
Db      937  CCATGCTTGGCCG 924

RESULT 61
US-11-217-529-3509/c
/ Sequence 3509, Application US/11217529
/ Publication No. US20060099612A1
/ GENERAL INFORMATION:
/ APPLICANT: SUNTORY LIMITED
/ APPLICANT: NAKAO, YOSHIHIRO
/ APPLICANT: NAKAMURA, NORIHIISA
/ APPLICANT: KODAMA, YUKIKO
/ APPLICANT: FUJIMURA, TOMOKO
/ APPLICANT: ASHIKARI, TOSHIHIKO
/ TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
/ FILE REFERENCE: S-38-285
/ CURRENT APPLICATION NUMBER: US/11/217,529
/ CURRENT FILING DATE: 2005-09-02
/ PRIOR APPLICATION NUMBER: US 10/932,182
/ PRIOR FILING DATE: 2004-09-02
/ NUMBER OF SEQ ID NOS: 197023
/ SOFTWARE: PatentIn version 3.3
/ SEQ ID NO 3509
/ LENGTH: 3093
/ TYPE: DNA
/ ORGANISM: Saccharomyces pastorianus
US-11-217-529-3509

Query Match          72.9%; Score 12.4; DB 7; Length 3093;
Best Local Similarity 92.9%; Pred. No. 2.3e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      4  CACGCTTGGCCGAG 17
Db      1452  CACGCTTGGCAGAG 1439

RESULT 62
US-10-527-552-5
/ Sequence 5, Application US/10527552
/ Publication No. US20060105424A1
/ GENERAL INFORMATION:
/ APPLICANT: Sandoz GmbH
/ TITLE OF INVENTION: Process for production of cephalosporin C
/ FILE REFERENCE: IB/G-32677A/BCK
/ CURRENT APPLICATION NUMBER: US/10/527,552
/ CURRENT FILING DATE: 2005-03-11
/ NUMBER OF SEQ ID NOS: 21
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 5
/ LENGTH: 16032
/ TYPE: DNA
/ ORGANISM: Acremonium chrysogenum
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (15909)..(15909)
/ OTHER INFORMATION: n is a, c, g, or t
US-10-527-552-5

Query Match          72.9%; Score 12.4; DB 6; Length 16032;
Best Local Similarity 92.9%; Pred. No. 2.4e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      4  CACGCTTGGCCGAG 17
Db      2109  CGCGCTTGGCCGAG 2122

RESULT 63
US-11-189-279-64
/ Sequence 64, Application US/11189279
/ Publication No. US20060115829A1
/ GENERAL INFORMATION:
/ APPLICANT: MAO, LI
/ APPLICANT: WANG, JIE
/ APPLICANT: LUO, WANG
/ TITLE OF INVENTION: A METHOD OF TREATING CANCER
/ FILE REFERENCE: UTXC:875US
/ CURRENT APPLICATION NUMBER: US/11/189,279
/ CURRENT FILING DATE: 2005-07-26
/ PRIOR APPLICATION NUMBER: 60/598,554
/ PRIOR FILING DATE: 2004-08-03
/ NUMBER OF SEQ ID NOS: 69
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 64
/ LENGTH: 118899
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-11-189-279-64

Query Match          72.9%; Score 12.4; DB 7; Length 118899;
Best Local Similarity 92.9%; Pred. No. 2.4e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3  CCACGCTTGGCCGA 16
Db      75324  CCACGCTTGGCTGA 75337

RESULT 64
US-11-259-950-11
/ Sequence 11, Application US/11259950
/ Publication No. US20060088524A1
/ GENERAL INFORMATION:
/ APPLICANT: Morrissey, James
/ APPLICANT: Pureza, Vincent
/ APPLICANT: Sliagar, Stephen G.
/ TITLE OF INVENTION: Tissue Factor Compositions and Methods
/ FILE REFERENCE: 46-04
/ CURRENT APPLICATION NUMBER: US/11/259,950
/ CURRENT FILING DATE: 2005-10-27
/ PRIOR APPLICATION NUMBER: US 60/622,737
/ PRIOR FILING DATE: 2004-10-27
/ PRIOR APPLICATION NUMBER: US 60/536,281
/ PRIOR FILING DATE: 2004-01-13
/ PRIOR APPLICATION NUMBER: US 10/465,789
/ PRIOR FILING DATE: 2003-06-18
/ PRIOR APPLICATION NUMBER: US 09/990,087
/ PRIOR FILING DATE: 2001-11-20
/ PRIOR APPLICATION NUMBER: US 60/252,233
/ PRIOR FILING DATE: 2000-11-20
/ PRIOR APPLICATION NUMBER: US 11/033,489
/ PRIOR FILING DATE: 2005-01-11
/ NUMBER OF SEQ ID NOS: 96
```



; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 11  
; LENGTH: 522  
; TYPE: DNA  
; ORGANISM: Artificial  
; FEATURE:  
; OTHER INFORMATION: Nucleotide sequence encoding MSP1D5D6  
US-11-259-950-11

Query Match 71.8%; Score 12.2; DB 7; Length 522;  
Best Local Similarity 82.4%; Pred. No. 2.7e+02;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GCCCAGCGCTTGGCGGAG 17  
||| ||||| |||||  
Db 293 GCCCAGCGCTTGGCGGAG 309

RESULT 65  
US-11-217-997-9  
; Sequence 9, Application US/11217997  
; Publication No. US20060111561A1  
; GENERAL INFORMATION:  
; APPLICANT: Valerie L. Gerlach  
; APPLICANT: Elma R. Fernandes  
; APPLICANT: Richard A. Shimkets  
; APPLICANT: Meera Patturajan  
; APPLICANT: Vladimir Y. Gusev  
; APPLICANT: Stacie (Casman) Navara  
; APPLICANT: Velizar T. Tchernev  
; APPLICANT: David W. Anderson  
; APPLICANT: Xiaojia (Sasha) Guo  
; APPLICANT: Luca Rastelli  
; APPLICANT: Mei Zhong  
; APPLICANT: Muralidhara Padigaru  
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODED THEREBY  
; FILE REFERENCE: Cura 551 CIP  
; CURRENT APPLICATION NUMBER: US/11/217,997  
; CURRENT FILING DATE: 2005-08-31  
; PRIOR APPLICATION NUMBER: 10/453,372  
; PRIOR FILING DATE: 2003-06-03  
; PRIOR APPLICATION NUMBER: 10/055,877  
; PRIOR FILING DATE: 2002-01-22  
; PRIOR APPLICATION NUMBER: 60/262,892  
; PRIOR FILING DATE: 2001-01-19  
; PRIOR APPLICATION NUMBER: 60/263,598  
; PRIOR FILING DATE: 2001-01-23  
; PRIOR APPLICATION NUMBER: 60/263,799  
; PRIOR FILING DATE: 2001-01-24  
; PRIOR APPLICATION NUMBER: 60/264,117  
; PRIOR FILING DATE: 2001-01-25  
; PRIOR APPLICATION NUMBER: 60/264,139  
; PRIOR FILING DATE: 2001-01-25  
; PRIOR APPLICATION NUMBER: 60/264,478  
; PRIOR FILING DATE: 2001-01-26  
; PRIOR APPLICATION NUMBER: 60/263,351  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: 60/272,870  
; PRIOR FILING DATE: 2001-03-02  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 62  
; SOFTWARE: CuraSeqlist version 0.1  
; SEQ ID NO 9  
; LENGTH: 522  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (29)..(514)  
US-11-217-997-9

Query Match 71.8%; Score 12.2; DB 7; Length 522;  
Best Local Similarity 82.4%; Pred. No. 2.7e+02;

Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 1 GCCCAGCGCTTGGCGGAG 17  
||| ||||| |||||  
Db 62 GCCCAGCGCTTGGCGGAG 78

RESULT 66  
US-11-217-997-35  
; Sequence 35, Application US/11217997  
; Publication No. US20060111561A1  
; GENERAL INFORMATION:  
; APPLICANT: Valerie L. Gerlach  
; APPLICANT: Elma R. Fernandes  
; APPLICANT: Richard A. Shimkets  
; APPLICANT: Meera Patturajan  
; APPLICANT: Vladimir Y. Gusev  
; APPLICANT: Stacie (Casman) Navara  
; APPLICANT: Velizar T. Tchernev  
; APPLICANT: David W. Anderson  
; APPLICANT: Xiaojia (Sasha) Guo  
; APPLICANT: Luca Rastelli  
; APPLICANT: Mei Zhong  
; APPLICANT: Muralidhara Padigaru  
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODED THEREBY  
; FILE REFERENCE: Cura 551 CIP  
; CURRENT APPLICATION NUMBER: US/11/217,997  
; CURRENT FILING DATE: 2005-08-31  
; PRIOR APPLICATION NUMBER: 10/453,372  
; PRIOR FILING DATE: 2003-06-03  
; PRIOR APPLICATION NUMBER: 10/055,877  
; PRIOR FILING DATE: 2002-01-22  
; PRIOR APPLICATION NUMBER: 60/262,892  
; PRIOR FILING DATE: 2001-01-19  
; PRIOR APPLICATION NUMBER: 60/263,598  
; PRIOR FILING DATE: 2001-01-23  
; PRIOR APPLICATION NUMBER: 60/263,799  
; PRIOR FILING DATE: 2001-01-24  
; PRIOR APPLICATION NUMBER: 60/264,117  
; PRIOR FILING DATE: 2001-01-25  
; PRIOR APPLICATION NUMBER: 60/264,139  
; PRIOR FILING DATE: 2001-01-25  
; PRIOR APPLICATION NUMBER: 60/264,478  
; PRIOR FILING DATE: 2001-01-26  
; PRIOR APPLICATION NUMBER: 60/263,351  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: 60/272,870  
; PRIOR FILING DATE: 2001-03-02  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 62  
; SOFTWARE: CuraSeqlist version 0.1  
; SEQ ID NO 35  
; LENGTH: 522  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (21)..(520)  
US-11-217-997-35

Query Match 71.8%; Score 12.2; DB 7; Length 522;  
Best Local Similarity 82.4%; Pred. No. 2.7e+02;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 1 GCCCAGCGCTTGGCGGAG 17  
||| ||||| |||||  
Db 62 GCCCAGCGCTTGGCGGAG 78

RESULT 67  
US-11-330-650-2/c  
; Sequence 2, Application US/11330650  
; Publication No. US2006094055A1

```
; GENERAL INFORMATION:
; APPLICANT: Ding, Shou-wei
; APPLICANT: Li, Hong-wei
; APPLICANT: Li, Wan-xiang
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: RNA Silencing in Animals as an Antiviral Defense
; FILE REFERENCE: 023070-124100US
; CURRENT APPLICATION NUMBER: US/11/330,650
; CURRENT FILING DATE: 2006-01-11
; PRIOR APPLICATION NUMBER: US/10/150,283
; PRIOR FILING DATE: 2002-05-15
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 540
; TYPE: RNA
; ORGANISM: Hepatitis C virus
; FEATURE:
; OTHER INFORMATION: F protein
US-11-330-650-2

Query Match          71.8%; Score 12.2; DB 7; Length 540;
Best Local Similarity 82.4%; Pred. No. 2.7e+02;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCCCAGCCTTGCCGAG 17
   ||||| ||||| |||||
Db 324 GCCCAGCTAGGCCGAG 308

RESULT 68
US-10-488-619-1465/c
; Sequence 1465, Application US/10488619
; Publication No. US2006009578A1
; GENERAL INFORMATION:
; APPLICANT: Greenlee, Winner and Sullivan, P.C.
; TITLE OF INVENTION: Human Mitochondrial DNA Polymorphisms, Haplogroups, Associations
; FILE REFERENCE: 98-01 WO
; CURRENT APPLICATION NUMBER: US/10/488,619
; CURRENT FILING DATE: 2004-03-01
; NUMBER OF SEQ ID NOS: 3040
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1465
; LENGTH: 562
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-488-619-1465

Query Match          71.8%; Score 12.2; DB 6; Length 562;
Best Local Similarity 82.4%; Pred. No. 2.7e+02;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCCCAGCCTTGCCGAG 17
   ||||| ||||| |||||
Db 209 GCCCAGCCTTGCCGGG 193

RESULT 69
US-11-259-950-5
; Sequence 5, Application US/11259950
; Publication No. US20060088524A1
; GENERAL INFORMATION:
; APPLICANT: Morrissey, James
; APPLICANT: Pureza, Vincent
; APPLICANT: Sligar, Stephen G.
; TITLE OF INVENTION: Tissue Factor Compositions and Methods
; FILE REFERENCE: 46-04
; CURRENT APPLICATION NUMBER: US/11/259,950
; CURRENT FILING DATE: 2005-10-27
; PRIOR APPLICATION NUMBER: US 60/622,737
; PRIOR FILING DATE: 2004-10-27
; PRIOR APPLICATION NUMBER: US 60/536,281
; PRIOR FILING DATE: 2004-01-13
```

```
; PRIOR FILING DATE: 2004-01-13
; PRIOR APPLICATION NUMBER: US 10/465,789
; PRIOR FILING DATE: 2003-06-18
; PRIOR APPLICATION NUMBER: US 09/990,087
; PRIOR FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: US 60/252,233
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 11/033,489
; PRIOR FILING DATE: 2005-01-11
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 5
; LENGTH: 619
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: Nucleotide sequence encoding MSP1 without His tag
US-11-259-950-5

Query Match          71.8%; Score 12.2; DB 7; Length 619;
Best Local Similarity 82.4%; Pred. No. 2.7e+02;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCCCAGCCTTGCCGAG 17
   ||||| ||||| |||||
Db 390 GCCCAGCCTTGCCGGG 406

RESULT 70
US-10-953-349-14927/c
; Sequence 14927, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 14927
; LENGTH: 633
; TYPE: DNA
; ORGANISM: Glycine max
US-10-953-349-14927

Query Match          71.8%; Score 12.2; DB 6; Length 633;
Best Local Similarity 82.4%; Pred. No. 2.7e+02;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCCCAGCCTTGCCGAG 17
   ||||| ||||| |||||
Db 97 GCCCAGCCTTGCCGAG 81

RESULT 71
US-11-259-950-3
; Sequence 3, Application US/11259950
; Publication No. US20060088524A1
; GENERAL INFORMATION:
; APPLICANT: Morrissey, James
; APPLICANT: Pureza, Vincent
; APPLICANT: Sligar, Stephen G.
; TITLE OF INVENTION: Tissue Factor Compositions and Methods
; FILE REFERENCE: 46-04
; CURRENT APPLICATION NUMBER: US/11/259,950
; CURRENT FILING DATE: 2005-10-27
; PRIOR APPLICATION NUMBER: US 60/622,737
; PRIOR FILING DATE: 2004-10-27
; PRIOR APPLICATION NUMBER: US 60/536,281
; PRIOR FILING DATE: 2004-01-13
; PRIOR APPLICATION NUMBER: US 10/465,789
```

; PRIOR FILING DATE: 2003-06-18  
; PRIOR APPLICATION NUMBER: US 09/990,087  
; PRIOR FILING DATE: 2001-11-20  
; PRIOR APPLICATION NUMBER: US 60/252,233  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 11/033,489  
; PRIOR FILING DATE: 2005-01-11  
; NUMBER OF SEQ ID NOS: 96  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 3  
; LENGTH: 654  
; TYPE: DNA  
; ORGANISM: Artificial  
; FEATURE:  
; OTHER INFORMATION: Nucleotide sequence encoding MSPI  
US-11-259-950-3

Query Match 71.8%; Score 12.2; DB 7; Length 654;  
Best Local Similarity 82.4%; Pred. No. 2.7e+02;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GCCCAGCGCTTGGCCGAG 17  
||| ||||| ||||| ||  
Db 425 GCCCAGCGCTTGGCCGCG 441

RESULT 72  
US-10-488-619-2847  
; Sequence 2847, Application US/10488619  
; Publication No. US20060099578A1  
; GENERAL INFORMATION:  
; APPLICANT: Greenlee, Winner and Sullivan, P.C.  
; TITLE OF INVENTION: Human Mitochondrial DNA Polymorphisms, Haplogroups, Associations  
; FILE REFERENCE: 98-01 WO  
; CURRENT APPLICATION NUMBER: US/10/488,619  
; CURRENT FILING DATE: 2004-03-01  
; NUMBER OF SEQ ID NOS: 3040  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2847  
; LENGTH: 655  
; TYPE: DNA  
; ORGANISM: Mus musculus  
US-10-488-619-2847

Query Match 71.8%; Score 12.2; DB 6; Length 655;  
Best Local Similarity 82.4%; Pred. No. 2.7e+02;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GCCCAGCGCTTGGCCGAG 17  
||| ||||| ||||| ||  
Db 483 GCCCAGCGCTTGGTCCAG 499

RESULT 73  
US-11-217-529-566  
; Sequence 566, Application US/11217529  
; Publication No. US20060099612A1  
; GENERAL INFORMATION:  
; APPLICANT: SUNTORY LIMITED  
; APPLICANT: NAKAO, YOSHIRO  
; APPLICANT: NAKAMURA, NORIHISA  
; APPLICANT: KODAMA, YUKIO  
; APPLICANT: FUJIMURA, TOMOKO  
; TITLE OF INVENTION: ASHIKARI, TOSHIHIKO  
; FILE REFERENCE: S-38-285  
; CURRENT APPLICATION NUMBER: US/11/217,529  
; CURRENT FILING DATE: 2005-09-02  
; PRIOR APPLICATION NUMBER: US 10/932,182  
; PRIOR FILING DATE: 2004-09-02  
; NUMBER OF SEQ ID NOS: 197023  
; SOFTWARE: PatentIn version 3.3

; SEQ ID NO 566  
; LENGTH: 681  
; TYPE: DNA  
; ORGANISM: Saccharomyces pastorianus  
US-11-217-529-566

Query Match 71.8%; Score 12.2; DB 7; Length 681;  
Best Local Similarity 82.4%; Pred. No. 2.7e+02;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GCCCAGCGCTTGGCCGAG 17  
||| ||||| ||||| ||  
Db 378 GCCCAGCGTTTGGCCAAG 394

RESULT 74  
US-11-259-950-1  
; Sequence 1, Application US/11259950  
; Publication No. US20060088524A1  
; GENERAL INFORMATION:  
; APPLICANT: Morrissey, James  
; APPLICANT: Pureza, Vincent  
; TITLE OF INVENTION: Tissue Factor Compositions and Methods  
; FILE REFERENCE: 46-04  
; CURRENT APPLICATION NUMBER: US/11/259,950  
; CURRENT FILING DATE: 2005-10-27  
; PRIOR APPLICATION NUMBER: US 60/622,737  
; PRIOR FILING DATE: 2004-10-27  
; PRIOR APPLICATION NUMBER: US 60/536,281  
; PRIOR FILING DATE: 2004-01-13  
; PRIOR APPLICATION NUMBER: US 10/465,789  
; PRIOR FILING DATE: 2003-06-18  
; PRIOR APPLICATION NUMBER: US 09/990,087  
; PRIOR FILING DATE: 2001-11-20  
; PRIOR APPLICATION NUMBER: US 60/252,233  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 11/033,489  
; PRIOR FILING DATE: 2005-01-11  
; NUMBER OF SEQ ID NOS: 96  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 1  
; LENGTH: 762  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-259-950-1

Query Match 71.8%; Score 12.2; DB 7; Length 762;  
Best Local Similarity 82.4%; Pred. No. 2.7e+02;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GCCCAGCGCTTGGCCGAG 17  
||| ||||| ||||| ||  
Db 535 GCCCAGCGCTTGGCCGCG 551

RESULT 75  
US-10-953-349-28486  
; Sequence 28486, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; FILE REFERENCE: ENCODED THEREBY  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 28486  
; LENGTH: 798  
; TYPE: DNA  
; ORGANISM: Triticum aestivum

US-10-953-349-28486

Query Match 71.8%; Score 12.2; DB 6; Length 798;  
Best Local Similarity 82.4%; Pred. No. 2.7e+02;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCCACGCTTGCCGAG 17  
| | | | | | | | | | | | | | | | | | | |  
Db 687 GTCCACGCTTGCCGAG 703

RESULT 76

US-10-196-749-487  
; Sequence 487, Application US/10196749  
; Publication No. US20060094864A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Chen, Jian  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Pan, James  
; APPLICANT: Smith, Victoria  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3430R1C340  
; CURRENT APPLICATION NUMBER: US/10/196,749  
; CURRENT FILING DATE: 2002-07-16  
; PRIOR APPLICATION NUMBER: 10/052586  
; PRIOR FILING DATE: 2002-01-15  
; PRIOR APPLICATION NUMBER: 60/059263  
; PRIOR FILING DATE: 1997-09-18  
; PRIOR APPLICATION NUMBER: 60/059266  
; PRIOR FILING DATE: 1997-09-18  
; PRIOR APPLICATION NUMBER: 60/062250  
; PRIOR FILING DATE: 1997-10-17  
; PRIOR APPLICATION NUMBER: 60/063120  
; PRIOR FILING DATE: 1997-10-24  
; PRIOR APPLICATION NUMBER: 60/063121  
; PRIOR FILING DATE: 1997-10-24  
; PRIOR APPLICATION NUMBER: 60/063486  
; PRIOR FILING DATE: 1997-10-21  
; PRIOR APPLICATION NUMBER: 60/063540  
; PRIOR FILING DATE: 1997-10-28  
; PRIOR APPLICATION NUMBER: 60/063541  
; PRIOR FILING DATE: 1997-10-28  
; PRIOR APPLICATION NUMBER: 60/063544  
; PRIOR FILING DATE: 1997-10-28  
; Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 612  
; SEQ ID NO 487  
; LENGTH: 843  
; TYPE: DNA  
; ORGANISM: Homo Sapien

US-10-196-749-487

Query Match 71.8%; Score 12.2; DB 6; Length 843;  
Best Local Similarity 82.4%; Pred. No. 2.7e+02;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCCACGCTTGCCGAG 17  
| | | | | | | | | | | | | | | | | | | |  
Db 767 GGCCACTCTTGCCGAG 783

RESULT 77

US-10-953-349-20319  
; Sequence 20319, Application US/10953349  
; Publication No. US20060107345A1

; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 20319  
; LENGTH: 868  
; TYPE: DNA  
; ORGANISM: Glycine max  
US-10-953-349-20319

Query Match 71.8%; Score 12.2; DB 6; Length 868;  
Best Local Similarity 82.4%; Pred. No. 2.8e+02;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCCACGCTTGCCGAG 17  
| | | | | | | | | | | | | | | | | | | |  
Db 519 GCCGCGCTTGCCGAG 535

RESULT 78

US-11-217-997-7  
; Sequence 7, Application US/11217997  
; Publication No. US20060111561A1  
; GENERAL INFORMATION:  
; APPLICANT: Valerie L. Gerlach  
; APPLICANT: Elma R. Fernandes  
; APPLICANT: Richard A. Shimkets  
; APPLICANT: Meera Patturajan  
; APPLICANT: Vladimir Y. Gusev  
; APPLICANT: Stacie (Casman) Navara  
; APPLICANT: Velizar T. Tchernev  
; APPLICANT: David W. Anderson  
; APPLICANT: Xiaojia (Sasha) Guo  
; APPLICANT: Luca Rastelli  
; APPLICANT: Mei Zhong  
; APPLICANT: Muralidhara Padigaru  
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODED THEREBY  
; FILE REFERENCE: Cura 551 CIP  
; CURRENT APPLICATION NUMBER: US/11/217,997  
; CURRENT FILING DATE: 2005-08-31  
; PRIOR APPLICATION NUMBER: 10/453,372  
; PRIOR FILING DATE: 2003-06-03  
; PRIOR APPLICATION NUMBER: 10/055,877  
; PRIOR FILING DATE: 2002-01-22  
; PRIOR APPLICATION NUMBER: 60/262,892  
; PRIOR FILING DATE: 2001-01-19  
; PRIOR APPLICATION NUMBER: 60/263,598  
; PRIOR FILING DATE: 2001-01-23  
; PRIOR APPLICATION NUMBER: 60/263,799  
; PRIOR FILING DATE: 2001-01-24  
; PRIOR APPLICATION NUMBER: 60/264,117  
; PRIOR FILING DATE: 2001-01-25  
; PRIOR APPLICATION NUMBER: 60/264,139  
; PRIOR FILING DATE: 2001-01-25  
; PRIOR APPLICATION NUMBER: 60/264,478  
; PRIOR FILING DATE: 2001-01-26  
; PRIOR APPLICATION NUMBER: 60/263,351  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: 60/272,870  
; PRIOR FILING DATE: 2001-03-02  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 62  
; SOFTWARE: CuraSequist version 0.1  
; SEQ ID NO 7  
; LENGTH: 877  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:

US-11-217-997-7

; NAME/KEY: CDS  
; LOCATION: (25)..(534)  
US-11-217-997-7

Query Match 71.8%; Score 12.2; DB 7; Length 877;  
Best Local Similarity 82.4%; Pred. No. 2.8e+02;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCCCAGCGCTGGCCGAG 17  
|||||  
Db 58 GCCCCCGCTGGCCGAG 74

## RESULT 79

US-10-953-349-28667/c  
; Sequence 28667, Application US/10953349  
; Publication No. US20060107345A1

## GENERAL INFORMATION:

; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 28667  
; LENGTH: 945  
; TYPE: DNA  
; ORGANISM: Triticum aestivum  
US-10-953-349-28667

Query Match 71.8%; Score 12.2; DB 6; Length 945;  
Best Local Similarity 82.4%; Pred. No. 2.8e+02;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCCCAGCGCTGGCCGAG 17  
|||||  
Db 374 GCCCGCGCTGGTCGAG 358

## RESULT 80

US-11-256-428-62  
; Sequence 62, Application US/11256428  
; Publication No. US20060095987A1

## GENERAL INFORMATION:

; APPLICANT: Niblett, Charles L.  
; TITLE OF INVENTION: Methods and Materials for Conferring Resistance to Pests and Pathogens of Plants  
; FILE REFERENCE: VEN-100  
; CURRENT APPLICATION NUMBER: US/11/256,428  
; CURRENT FILING DATE: 2005-10-21  
; NUMBER OF SEQ ID NOS: 69  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 62  
; LENGTH: 949  
; TYPE: DNA  
; ORGANISM: Artificial  
; FEATURE:  
; OTHER INFORMATION: Heterodera glycines rDNA  
US-11-256-428-62

Query Match 71.8%; Score 12.2; DB 7; Length 949;  
Best Local Similarity 82.4%; Pred. No. 2.8e+02;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCCCAGCGCTGGCCGAG 17  
|||||  
Db 468 GCCCGCTGGCCTAG 484

## RESULT 81

US-10-953-349-27232

; Sequence 27232, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:

; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 27232  
; LENGTH: 993  
; TYPE: DNA  
; ORGANISM: Triticum aestivum  
US-10-953-349-27232

Query Match 71.8%; Score 12.2; DB 6; Length 993;  
Best Local Similarity 82.4%; Pred. No. 2.8e+02;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCCCAGCGCTGGCCGAG 17  
|||||  
Db 342 GCCCCAGCGCTGGCCGAG 358

## RESULT 82

US-10-953-349-3387  
; Sequence 3387, Application US/10953349  
; Publication No. US20060107345A1

## GENERAL INFORMATION:

; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 3387  
; LENGTH: 1038  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-10-953-349-3387

Query Match 71.8%; Score 12.2; DB 6; Length 1038;  
Best Local Similarity 82.4%; Pred. No. 2.8e+02;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCCCAGCGCTGGCCGAG 17  
|||||  
Db 699 GCCTACGCTGGCTAAG 715

## RESULT 83

US-10-953-349-35759/c  
; Sequence 35759, Application US/10953349  
; Publication No. US20060107345A1

## GENERAL INFORMATION:

; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 35759  
; LENGTH: 1233  
; TYPE: DNA  
; ORGANISM: Zea mays subsp. mays  
US-10-953-349-35759

```
Query Match          71.8%; Score 12.2; DB 6; Length 1233;
Best Local Similarity 82.4%; Pred. No. 2.8e+02;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCCCAGCGTTGGCCGAG 17
Db 774 GCCCCCGGAGGCCGAG 758

RESULT 84
US-11-254-252-1/C
; Sequence 1, Application US/11254252
; Publication No. US20060110755A1
; GENERAL INFORMATION:
; APPLICANT: Duke, Richard C.
; APPLICANT: Franzusoff, Alex
; APPLICANT: Haller, Aurelia
; APPLICANT: King, Thomas H.
; TITLE OF INVENTION: YEAST-BASED THERAPEUTIC FOR CHRONIC HEPATITIS C INFECTION
; FILE REFERENCE: 3923-12
; CURRENT APPLICATION NUMBER: US/11/254,252
; CURRENT FILING DATE: 2005-10-18
; PRIOR APPLICATION NUMBER: US 60/620,158
; PRIOR FILING DATE: 2004-10-18
; PRIOR APPLICATION NUMBER: US 10/738,646
; PRIOR FILING DATE: 2003-12-16
; PRIOR APPLICATION NUMBER: US 60/434,163
; PRIOR FILING DATE: 2002-12-16
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1
; LENGTH: 1233
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: recombinant fusion protein construct
US-11-254-252-1

Query Match          71.8%; Score 12.2; DB 7; Length 1233;
Best Local Similarity 82.4%; Pred. No. 2.8e+02;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCCCAGCGTTGGCCGAG 17
Db 1128 GCCCCAGCTAGGCCGAG 1112

RESULT 85
US-11-259-950-7
; Sequence 7, Application US/11259950
; Publication No. US20060088524A1
; GENERAL INFORMATION:
; APPLICANT: Morrissey, James
; APPLICANT: Pureza, Vincent
; APPLICANT: Sligar, Stephen G.
; TITLE OF INVENTION: Tissue Factor Compositions and Methods
; FILE REFERENCE: 46-04
; CURRENT APPLICATION NUMBER: US/11/259,950
; CURRENT FILING DATE: 2005-10-27
; PRIOR APPLICATION NUMBER: US 60/622,737
; PRIOR FILING DATE: 2004-10-27
; PRIOR APPLICATION NUMBER: US 60/536,281
; PRIOR FILING DATE: 2004-01-13
; PRIOR APPLICATION NUMBER: US 10/465,789
; PRIOR FILING DATE: 2003-06-18
; PRIOR APPLICATION NUMBER: US 09/990,087
; PRIOR FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: US 60/252,233
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 11/033,489
; PRIOR FILING DATE: 2005-01-11
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1
; LENGTH: 1233
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: Nucleotide sequence encoding MSP2TEV
US-11-259-950-7

Query Match          71.8%; Score 12.2; DB 7; Length 1275;
Best Local Similarity 82.4%; Pred. No. 2.8e+02;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCCCAGCGTTGGCCGAG 17
Db 452 GCCAGCGCTTGCCGCG 468

RESULT 87
US-11-259-950-9
; Sequence 9, Application US/11259950
; Publication No. US20060088524A1
; GENERAL INFORMATION:
; APPLICANT: Morrissey, James
; APPLICANT: Pureza, Vincent
; APPLICANT: Sligar, Stephen G.
; TITLE OF INVENTION: Tissue Factor Compositions and Methods
; FILE REFERENCE: 46-04
; CURRENT APPLICATION NUMBER: US/11/259,950
; CURRENT FILING DATE: 2005-10-27
; PRIOR APPLICATION NUMBER: US 60/622,737
```

```
; SEQ ID NO 7
; LENGTH: 1260
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: Nucleotide sequence encoding MSP2 with short linker
US-11-259-950-7

Query Match          71.8%; Score 12.2; DB 7; Length 1260;
Best Local Similarity 82.4%; Pred. No. 2.8e+02;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCCCAGCGTTGGCCGAG 17
Db 425 GCCAGCGCTTGCCGCG 441

RESULT 86
US-11-259-950-77
; Sequence 77, Application US/11259950
; Publication No. US20060088524A1
; GENERAL INFORMATION:
; APPLICANT: Morrissey, James
; APPLICANT: Pureza, Vincent
; APPLICANT: Sligar, Stephen G.
; TITLE OF INVENTION: Tissue Factor Compositions and Methods
; FILE REFERENCE: 46-04
; CURRENT APPLICATION NUMBER: US/11/259,950
; CURRENT FILING DATE: 2005-10-27
; PRIOR APPLICATION NUMBER: US 60/622,737
; PRIOR FILING DATE: 2004-10-27
; PRIOR APPLICATION NUMBER: US 60/536,281
; PRIOR FILING DATE: 2004-01-13
; PRIOR APPLICATION NUMBER: US 10/465,789
; PRIOR FILING DATE: 2003-06-18
; PRIOR APPLICATION NUMBER: US 09/990,087
; PRIOR FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: US 60/252,233
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 11/033,489
; PRIOR FILING DATE: 2005-01-11
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 77
; LENGTH: 1275
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: Nucleotide sequence encoding MSP2TEV
US-11-259-950-77

Query Match          71.8%; Score 12.2; DB 7; Length 1275;
Best Local Similarity 82.4%; Pred. No. 2.8e+02;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCCCAGCGTTGGCCGAG 17
Db 452 GCCAGCGCTTGCCGCG 468

RESULT 87
US-11-259-950-9
; Sequence 9, Application US/11259950
; Publication No. US20060088524A1
; GENERAL INFORMATION:
; APPLICANT: Morrissey, James
; APPLICANT: Pureza, Vincent
; APPLICANT: Sligar, Stephen G.
; TITLE OF INVENTION: Tissue Factor Compositions and Methods
; FILE REFERENCE: 46-04
; CURRENT APPLICATION NUMBER: US/11/259,950
; CURRENT FILING DATE: 2005-10-27
; PRIOR APPLICATION NUMBER: US 60/622,737
```

; PRIOR FILING DATE: 2004-10-27  
; PRIOR APPLICATION NUMBER: US 60/536,281  
; PRIOR FILING DATE: 2004-01-13  
; PRIOR APPLICATION NUMBER: US 10/465,789  
; PRIOR FILING DATE: 2003-06-18  
; PRIOR APPLICATION NUMBER: US 09/990,087  
; PRIOR FILING DATE: 2001-11-20  
; PRIOR APPLICATION NUMBER: US 60/252,233  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 11/033,489  
; PRIOR FILING DATE: 2005-01-11  
; NUMBER OF SEQ ID NOS: 96  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 9  
; LENGTH: 1282  
; TYPE: DNA  
; ORGANISM: Artificial

; FEATURE:  
; OTHER INFORMATION: Nucleotide sequence encoding MSP2L (with long linker)

US-11-259-950-9

Query Match 71.8%; Score 12.2; DB 7; Length 1282;  
Best Local Similarity 82.4%; Pred. No. 2.8e+02;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GCCCAGCGCTGGCCGAG 17  
||| |||||  
Db 423 GCCAGCGCTGGCCGCG 439

## RESULT 88

US-10-953-349-35397/c  
; Sequence 35397, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: ENCODED THERBY  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 35397  
; LENGTH: 1295  
; TYPE: DNA  
; ORGANISM: Zea mays subsp. mays

US-10-953-349-35397

Query Match 71.8%; Score 12.2; DB 6; Length 1295;  
Best Local Similarity 82.4%; Pred. No. 2.8e+02;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GCCCAGCGCTGGCCGAG 17  
||| |||||  
Db 247 GCCCGCGCTGGCCGAG 231

## RESULT 89

US-10-953-349-27665/c  
; Sequence 27665, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: ENCODED THERBY  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 27665  
; LENGTH: 1378

; TYPE: DNA  
; ORGANISM: Triticum aestivum  
US-10-953-349-27665

Query Match 71.8%; Score 12.2; DB 6; Length 1378;  
Best Local Similarity 82.4%; Pred. No. 2.8e+02;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GCCCAGCGCTGGCCGAG 17  
||| |||||  
Db 940 GCCCGCGCTGGCCGCG 924

## RESULT 90

US-10-953-349-28083  
; Sequence 28083, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: ENCODED THERBY  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 28083  
; LENGTH: 1409  
; TYPE: DNA  
; ORGANISM: Triticum aestivum  
US-10-953-349-28083

Query Match 71.8%; Score 12.2; DB 6; Length 1409;  
Best Local Similarity 82.4%; Pred. No. 2.8e+02;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GCCCAGCGCTGGCCGAG 17  
||| |||||  
Db 1340 GCCCATGATTGGCCGTG 1356

## RESULT 91

US-10-953-349-35361  
; Sequence 35361, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: ENCODED THERBY  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 35361  
; LENGTH: 1605  
; TYPE: DNA  
; ORGANISM: Zea mays subsp. mays  
US-10-953-349-35361

Query Match 71.8%; Score 12.2; DB 6; Length 1605;  
Best Local Similarity 82.4%; Pred. No. 2.8e+02;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GCCCAGCGCTGGCCGAG 17  
||| |||||  
Db 234 GCCCGCGCTTCGCCGTG 250

## RESULT 92

US-10-505-928-653  
; Sequence 653, Application US/10505928  
; Publication No. US2006008532A1

; GENERAL INFORMATION:  
; APPLICANT: Ludwig Institute for Cancer Research et al.  
; TITLE OF INVENTION: LYMPHATIC ENDOTHELIAL GENES  
; FILE REFERENCE: 28967/39178  
; CURRENT APPLICATION NUMBER: US/10/505,928  
; CURRENT FILING DATE: 2004-08-27  
; PRIOR APPLICATION NUMBER: US 60/363,019  
; PRIOR FILING DATE: 2002-03-07  
; NUMBER OF SEQ ID NOS: 866  
; SOFTWARE: PatentIn 3.2  
; SEQ ID NO 653  
; LENGTH: 1645  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-505-928-653

Query Match 71.8%; Score 12.2; DB 6; Length 1645;  
Best Local Similarity 82.4%; Pred. No. 2.8e+02;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCCCAGCGTTGGCCGAG 17  
||| ||||| |||  
Db 1569 GCCTACGCTTGGCCAG 1585

## RESULT 93

US-11-293-697-981  
; Sequence 981, Application US/11293697  
; Publication No. US20060105376A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: Novel full length cDNA  
; FILE REFERENCE: H1-A0106  
; CURRENT APPLICATION NUMBER: US/11/293,697  
; CURRENT FILING DATE: 2005-12-05  
; PRIOR APPLICATION NUMBER: US/10/108,260  
; PRIOR FILING DATE: 2002-03-28  
; NUMBER OF SEQ ID NOS: 5458  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 981  
; LENGTH: 1736  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-293-697-981

Query Match 71.8%; Score 12.2; DB 7; Length 1736;  
Best Local Similarity 82.4%; Pred. No. 2.8e+02;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCCCAGCGTTGGCCGAG 17  
||| ||||| |||||  
Db 620 GTCCAGCGTGGGCTAG 636

## RESULT 94

US-10-953-349-34645/c  
; Sequence 34645, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 34645  
; LENGTH: 1859  
; TYPE: DNA  
; ORGANISM: Zea mays subsp. mays  
US-10-953-349-34645

Query Match 71.8%; Score 12.2; DB 6; Length 1859;  
Best Local Similarity 82.4%; Pred. No. 2.8e+02;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCCCAGCGTTGGCCGAG 17  
||| ||||| |||||  
Db 1375 GCTAGCGTTGGCCGAG 1359

## RESULT 95

US-11-254-252-13/c  
; Sequence 13, Application US/11254252  
; Publication No. US20060110755A1  
; GENERAL INFORMATION:  
; APPLICANT: Duke, Richard C.  
; APPLICANT: Franzusoff, Alex  
; APPLICANT: Haller, Aurelia  
; APPLICANT: King, Thomas H  
; TITLE OF INVENTION: YEAST-BASED THERAPEUTIC FOR CHRONIC HEPATITIS C INFECTION  
; FILE REFERENCE: 3923-12  
; CURRENT APPLICATION NUMBER: US/11/254,252  
; CURRENT FILING DATE: 2005-10-18  
; PRIOR APPLICATION NUMBER: US 60/620,158  
; PRIOR FILING DATE: 2004-10-18  
; PRIOR APPLICATION NUMBER: US 10/738,646  
; PRIOR FILING DATE: 2003-12-16  
; PRIOR APPLICATION NUMBER: US 60/434,163  
; PRIOR FILING DATE: 2002-12-16  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 13  
; LENGTH: 1908  
; TYPE: DNA  
; ORGANISM: Artificial  
; FEATURE:  
; OTHER INFORMATION: recombinant fusion protein construct  
US-11-254-252-13

Query Match 71.8%; Score 12.2; DB 7; Length 1908;  
Best Local Similarity 82.4%; Pred. No. 2.9e+02;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCCCAGCGTTGGCCGAG 17  
||| ||| |||||  
Db 339 GCCCAGCTAGGCCGAG 323

## RESULT 96

US-11-217-529-79927/c  
; Sequence 79927, Application US/11217529  
; Publication No. US20060099612A1  
; GENERAL INFORMATION:  
; APPLICANT: SUNTORY LIMITED  
; APPLICANT: NAKAO, YOSHIHIRO  
; APPLICANT: NAKAMURA, NORIHISA  
; APPLICANT: KODAMA, YUKIKO  
; APPLICANT: FUJIMURA, TOMOKO  
; APPLICANT: ASHIKARI, TOSHIHIKO  
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS  
; FILE REFERENCE: S-38-285  
; CURRENT APPLICATION NUMBER: US/11/217,529  
; CURRENT FILING DATE: 2005-09-02  
; PRIOR APPLICATION NUMBER: US 10/932,182  
; PRIOR FILING DATE: 2004-09-02  
; NUMBER OF SEQ ID NOS: 197023  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 79927  
; LENGTH: 1941  
; TYPE: DNA  
; ORGANISM: Saccharomyces pastorianus  
US-11-217-529-79927

Query Match 71.8%; Score 12.2; DB 7; Length 1941;



Best Local Similarity 82.4%; Pred. No. 2.9e+02;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 1 GCCCAGCGCTTGCCGAG 17  
Db 408 GCCCAGCGCTTGACGAG 392

## RESULT 97

US-11-293-697-2383  
; Sequence 2383, Application US/11293697  
; Publication No. US20060105376A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: Novel full length cDNA  
; FILE REFERENCE: HI-A0106  
; CURRENT APPLICATION NUMBER: US/11/293,697  
; CURRENT FILING DATE: 2005-12-05  
; PRIOR APPLICATION NUMBER: US/10/108,260  
; PRIOR FILING DATE: 2002-03-28  
; NUMBER OF SEQ ID NOS: 5458  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2383  
; LENGTH: 2128  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-293-697-2383

Query Match 71.8%; Score 12.2; DB 7; Length 2128;  
Best Local Similarity 82.4%; Pred. No. 2.9e+02;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 1 GCCCAGCGCTTGCCGAG 17  
Db 800 GCCCAGCGCTTGCCGAG 816

## RESULT 98

US-11-293-697-1876/c  
; Sequence 1876, Application US/11293697  
; Publication No. US20060105376A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: Novel full length cDNA  
; FILE REFERENCE: HI-A0106  
; CURRENT APPLICATION NUMBER: US/11/293,697  
; CURRENT FILING DATE: 2005-12-05  
; PRIOR APPLICATION NUMBER: US/10/108,260  
; PRIOR FILING DATE: 2002-03-28  
; NUMBER OF SEQ ID NOS: 5458  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1876  
; LENGTH: 2182  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-293-697-1876

Query Match 71.8%; Score 12.2; DB 7; Length 2182;  
Best Local Similarity 82.4%; Pred. No. 2.9e+02;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 1 GCCCAGCGCTTGCCGAG 17  
Db 1523 GCCCAGCGCTTGCCGAG 1507

## RESULT 99

US-11-293-697-40  
; Sequence 40, Application US/11293697  
; Publication No. US20060105376A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: Novel full length cDNA

; FILE REFERENCE: HI-A0106  
; CURRENT APPLICATION NUMBER: US/11/293,697  
; CURRENT FILING DATE: 2005-12-05  
; PRIOR APPLICATION NUMBER: US/10/108,260  
; PRIOR FILING DATE: 2002-03-28  
; NUMBER OF SEQ ID NOS: 5458  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 40  
; LENGTH: 2208  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-293-697-40

Query Match 71.8%; Score 12.2; DB 7; Length 2208;  
Best Local Similarity 82.4%; Pred. No. 2.9e+02;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 1 GCCCAGCGCTTGCCGAG 17  
Db 659 GCCCAGCGCTTGCTGAG 675

## RESULT 100

US-11-293-697-40/c  
; Sequence 40, Application US/11293697  
; Publication No. US20060105376A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: Novel full length cDNA  
; FILE REFERENCE: HI-A0106  
; CURRENT APPLICATION NUMBER: US/11/293,697  
; CURRENT FILING DATE: 2005-12-05  
; PRIOR APPLICATION NUMBER: US/10/108,260  
; PRIOR FILING DATE: 2002-03-28  
; NUMBER OF SEQ ID NOS: 5458  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 40  
; LENGTH: 2208  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-293-697-40

Query Match 71.8%; Score 12.2; DB 7; Length 2208;  
Best Local Similarity 82.4%; Pred. No. 2.9e+02;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 1 GCCCAGCGCTTGCCGAG 17  
Db 1770 GCCCAGCGCATGCCGCGG 1754

Search completed: June 10, 2006, 15:37:52  
Job time : 14.2167 secs

**THIS PAGE IS BLANK**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2006, 15:18:44 ; Search time 518.067 Seconds  
(without alignments)  
2345.260 Million cell updates/sec

Title: US-10-600-816-32

Perfect score: 19

Sequence: 1 ctttacttcattgctttg 19

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 6366136 seqs, 31973710525 residues

Total number of hits satisfying chosen parameters: 12732272

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

Database :

GenEmbl.\*

1: gb\_env.\*

2: gb\_pat.\*

3: gb\_ph.\*

4: gb\_pl.\*

5: gb\_pr.\*

6: gb\_ro.\*

7: gb\_sts.\*

8: gb\_sy.\*

9: gb\_un.\*

10: gb\_vi.\*

11: gb\_ov.\*

12: gb\_htg.\*

13: gb\_in.\*

14: gb\_om.\*

15: gb\_ba.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	19	100.0	603	2	BD233463 Human pro
C 2	19	100.0	1212	2	BD209699 Compositi
C 3	19	100.0	1212	2	AR341505 Sequence
C 4	19	100.0	1619	2	AX078375 Sequence
C 5	19	100.0	1718	2	BD233473 Human pro
C 6	19	100.0	1826	5	AK172760 Homo sapi
C 7	19	100.0	2290	2	C0723177 Sequence
C 8	19	100.0	2295	5	BC003665 Homo sapi
C 9	19	100.0	2297	5	AF506289 Homo sapi
C 10	19	100.0	2302	2	AX930411 Sequence
C 11	19	100.0	2302	5	AF095448 Homo sapi
C 12	19	100.0	2446	2	BD156680 Primer fo
C 13	19	100.0	2446	2	AX877483 Sequence
C 14	19	100.0	2446	5	AK001761 Homo sapi
C 15	19	100.0	2456	2	C0981495 Sequence
C 16	19	100.0	2456	2	DD210040 Methods o
C 17	19	100.0	2456	2	AX549168 Sequence
C 18	19	100.0	3057	5	AK122672 Homo sapi

AX188348 Sequence  
C0894732 Sequence  
AC007688 Homo sapi  
AL954230 Pan trogl  
AF064865 Homo sapi  
AC087481 Homo sapi  
AL954231 Pan trogl  
AC073220 Homo sapi  
AL954232 Pan trogl  
AC099102 Rattus no  
CT027583 Danio rer  
AL163281 Homo sapi  
DQ271038 Mustela v  
Z72916 S.cerevisia  
Z72915 S.cerevisia  
AY800241 Homo sapi  
AC157840 Aotus nan  
Continuation (68 o  
Continuation (150  
Continuation (25 o  
AP001818 Homo sapi  
AC150029 Sorex ara  
AC167325 Loxodonta  
AL121825 Human DNA  
AC011354 Homo sapi  
AC011334 Homo sapi  
AP006464 Oryza sat  
AC091155 Homo sapi  
AC074120 Homo sapi  
AP003856 Oryza sat  
AP003556 Homo sapi  
AC173916 Ateleterix  
AP003021 Oryza sat  
AC015570 Homo sapi  
AC021588 Homo sapi  
AC124725 Mus muscu  
AP003174 Homo sapi  
AC127007 Rattus no  
AC164008 Mus muscu  
AC148231 Didelphis  
AL845354 Mouse DNA  
AL512635 Human DNA  
AC011949 Homo sapi  
AC022743 Homo sapi  
AC097585 Sus scrof  
AC146837 Canis fam  
AC011947 Homo sapi  
AC069476 Homo sapi  
AC132501 Rattus no  
AC161349 Mus muscu  
AC144518 Pan trogl  
AC151368 Aotus nan  
AC097629 Sus scrof  
AC122492 Mus muscu  
AC097625 Homo sapi  
AP001797 Homo sapi  
AC122492 Mus muscu  
AC102703 Mus muscu  
AC151118 Mus muscu  
AC146886 Callithri  
AC113245 Mus muscu  
AC146883 Callithri  
AC147845 Saimiri b  
AC092434 Homo sapi  
AC112307 Rattus no  
AC096378 Rattus no  
AC095186 Rattus no  
AC137204 Rattus no  
AC108958 Rattus no  
AC116211 Rattus no  
AB219542 Campyloba  
AC172882 Brassica  
Continuation (2 of  
Continuation (2 of













822	15.8	83.2	171376	12	AC171476	AC171476 Bos tauru	895	15.8	83.2	185297	6	AC102250	Mus muscu
823	15.8	83.2	171648	12	AC171403	Mus muscu	896	15.8	83.2	185494	12	AC169055	Bos tauru
824	15.8	83.2	171716	5	AC034187	Homo sapi	897	15.8	83.2	185617	6	AC140235	Mus muscu
825	15.8	83.2	171834	12	AC174213	Bos tauru	898	15.8	83.2	186218	5	AP006248	Homo sapi
826	15.8	83.2	171937	5	AC008442	Homo sapi	899	15.8	83.2	186311	6	AC166161	Mus muscu
827	15.8	83.2	172300	5	AC010254	Homo sapi	900	15.8	83.2	186432	13	AC099034	Drosophil
828	15.8	83.2	172493	12	AC146888	Bos tauru	901	15.8	83.2	186607	12	AC007772	Drosophil
829	15.8	83.2	172543	5	AP004287	Homo sapi	902	15.8	83.2	186719	12	AC147574	Homo sapi
830	15.8	83.2	172579	5	AC008064	Homo sapi	903	15.8	83.2	186747	12	AC018605	Homo sapi
831	15.8	83.2	172942	5	AC123537	Macaca mu	904	15.8	83.2	187487	12	CT573220	Danio rer
832	15.8	83.2	173312	12	AC162865	Mus muscu	905	15.8	83.2	187677	12	CT573220	Danio rer
833	15.8	83.2	173351	6	AC127562	Mus muscu	906	15.8	83.2	187885	12	AC162985	Bos tauru
834	15.8	83.2	173534	5	AC007920	Homo sapi	907	15.8	83.2	188483	12	AC176780	Strongylo
835	15.8	83.2	173819	12	AC161014	Pan trogl	908	15.8	83.2	188665	12	CT025849	Danio rer
836	15.8	83.2	173821	5	AC020599	Homo sapi	909	15.8	83.2	188864	12	AC135951	Macaca mu
837	15.8	83.2	174032	12	AC128150	Rattus no	910	15.8	83.2	189033	6	AC117690	Mus muscu
838	15.8	83.2	174037	5	AC009995	Homo sapi	911	15.8	83.2	189139	12	AC122947	Rattus no
839	15.8	83.2	174267	6	AC132327	Mus muscu	912	15.8	83.2	189500	6	AC157656	Mus muscu
840	15.8	83.2	174333	5	AL356454	Human DNA	913	15.8	83.2	189893	4	AP005298	Oryza sat
841	15.8	83.2	174407	5	AC138725	Cercopith	914	15.8	83.2	189941	12	AC128854	Rattus no
842	15.8	83.2	174500	12	AC113089	Mus muscu	915	15.8	83.2	190673	12	AC166022	Oryctolag
843	15.8	83.2	174507	6	AC132955	Mus muscu	916	15.8	83.2	190679	12	AC147747	Homo sapi
844	15.8	83.2	174640	12	AC117840	Rattus no	917	15.8	83.2	190680	5	AC147575	Homo sapi
845	15.8	83.2	174962	5	AC145145	Homo sapi	918	15.8	83.2	190702	11	BX119962	Zebrafish
846	15.8	83.2	175291	12	AC177990	Strongylo	919	15.8	83.2	190712	6	AC100382	Mus muscu
847	15.8	83.2	175706	12	AC013798	Homo sapi	920	15.8	83.2	191462	12	AP005781	Oryza sat
848	15.8	83.2	175838	6	AC122495	Mus muscu	921	15.8	83.2	191544	12	AC138855	Homo sapi
849	15.8	83.2	175992	5	AC016943	Homo sapi	922	15.8	83.2	191662	6	AL772296	Mouse DNA
850	15.8	83.2	176018	5	AC093878	Homo sapi	923	15.8	83.2	192026	6	AC157660	Mus muscu
851	15.8	83.2	176134	12	AC181356	Strongylo	924	15.8	83.2	192044	5	AL590439	Human DNA
852	15.8	83.2	176222	11	AL929306	Zebrafish	925	15.8	83.2	192356	5	AC162122	Colobus g
853	15.8	83.2	177191	5	AC145146	Homo sapi	926	15.8	83.2	192499	12	AC170431	Bos tauru
854	15.8	83.2	177350	6	AC123743	Mus muscu	927	15.8	83.2	192699	6	AC133597	Mus muscu
855	15.8	83.2	177535	12	AC109914	Bos tauru	928	15.8	83.2	192944	12	AC146902	Callicebu
856	15.8	83.2	177917	6	AC115045	Mus muscu	929	15.8	83.2	192961	5	AP002770	Homo sapi
857	15.8	83.2	178262	12	AC167516	Sorex ara	930	15.8	83.2	193047	5	AC147069	Pan trogl
858	15.8	83.2	178273	6	AL954341	Mouse DNA	931	15.8	83.2	193170	6	AC115709	Mus muscu
859	15.8	83.2	178605	12	AC160869	Orolemur	932	15.8	83.2	193191	6	AC149276	Mus muscu
860	15.8	83.2	178781	6	AC165424	Mus muscu	933	15.8	83.2	193219	6	AL807816	Mouse DNA
861	15.8	83.2	178965	6	AL844144	Mouse DNA	934	15.8	83.2	193277	6	AC104324	Mus muscu
862	15.8	83.2	179125	13	AC009392	Drosophil	935	15.8	83.2	193643	5	AC018644	Homo sapi
863	15.8	83.2	179476	11	BX294126	Zebrafish	936	15.8	83.2	194020	12	AC101988	Mus muscu
864	15.8	83.2	179532	12	AC148884	Orolemur	937	15.8	83.2	194095	5	AL929401	Human DNA
865	15.8	83.2	179723	5	AC087858	Homo sapi	938	15.8	83.2	194439	12	AC109022	Rattus no
866	15.8	83.2	179777	12	AC067879	Homo sapi	939	15.8	83.2	194476	12	AC147745	Homo sapi
867	15.8	83.2	180125	6	AC159186	Mus muscu	940	15.8	83.2	194718	12	AP006254	Homo sapi
868	15.8	83.2	180152	5	BS000052	Pan trogl	941	15.8	83.2	194748	12	BX890638	Mouse DNA
869	15.8	83.2	180201	6	AC131736	Mus muscu	942	15.8	83.2	195039	6	BX649621	Mouse DNA
870	15.8	83.2	180510	12	AC181368	Strongylo	943	15.8	83.2	195169	5	AC017109	Homo sapi
871	15.8	83.2	180651	4	AP005745	Oryza sat	944	15.8	83.2	195646	5	AC165377	Nonascul
872	15.8	83.2	180717	12	AC181132	Strongylo	945	15.8	83.2	195833	12	AC168568	Strongylo
873	15.8	83.2	180889	12	AC079986	Homo sapi	946	15.8	83.2	196422	12	AC147601	Homo sapi
874	15.8	83.2	181233	12	AC149832	Zea mays	947	15.8	83.2	196518	5	AC116154	Homo sapi
875	15.8	83.2	181308	5	AL353148	Human DNA	948	15.8	83.2	196567	6	AL928855	Mouse DNA
876	15.8	83.2	181847	6	AL512583	Mouse DNA	949	15.8	83.2	196860	6	AL168279	Mus muscu
877	15.8	83.2	182038	5	AC073042	Homo sapi	950	15.8	83.2	197031	12	AC019121	Homo sapi
878	15.8	83.2	182340	12	AC024417	Homo sapi	951	15.8	83.2	197292	12	AC182010	Echinops
879	15.8	83.2	182431	6	AC151558	Dasypus n	952	15.8	83.2	197316	6	AC134899	Mus muscu
880	15.8	83.2	182520	12	AL627392	Mouse DNA	953	15.8	83.2	197430	12	AC140273	Mus muscu
881	15.8	83.2	182599	12	AC022897	Homo sapi	954	15.8	83.2	197565	11	CR450724	Zebrafish
882	15.8	83.2	182658	6	AC164549	Mus muscu	955	15.8	83.2	197602	12	AC175711	Pan trogl
883	15.8	83.2	182912	5	AC068060	Homo sapi	956	15.8	83.2	197646	6	AC136376	Mus muscu
884	15.8	83.2	182931	5	AC068556	Homo sapi	957	15.8	83.2	197648	6	AC164400	Mus muscu
885	15.8	83.2	183034	6	AC154615	Mus muscu	958	15.8	83.2	197693	12	AC146832	Bos tauru
886	15.8	83.2	183295	12	AC159805	Glycine m	959	15.8	83.2	198006	6	AC119210	Mus muscu
887	15.8	83.2	183355	12	AC178035	Strongylo	960	15.8	83.2	198021	12	AC152620	Bos tauru
888	15.8	83.2	183487	12	AC068828	Homo sapi	961	15.8	83.2	198307	6	AC132432	Mus muscu
889	15.8	83.2	183577	12	AC011164	Homo sapi	962	15.8	83.2	198318	12	CT486000	Mus muscu
890	15.8	83.2	183850	12	AC172889	Phytophch	963	15.8	83.2	198396	5	AC010355	Homo sapi
891	15.8	83.2	184436	6	AL671870	Mouse DNA	964	15.8	83.2	198440	12	CR925729	Danio rer
892	15.8	83.2	184700	6	AL671870	Mouse DNA	965	15.8	83.2	198575	5	AC007495	Homo sapi
893	15.8	83.2	185091	12	AC015835	Homo sapi	966	15.8	83.2	198635	12	AC121182	Rattus no
894	15.8	83.2	185121	12	AC079806	Homo sapi	967	15.8	83.2	198693	12	AC166721	Bos tauru

```

c 968 15.8 83.2 198880 12 AC118266 Mus muscu
c 969 15.8 83.2 198885 12 CT033785 Mus muscu
c 970 15.8 83.2 199014 12 AC073849 Homo sapi
c 971 15.8 83.2 199191 6 AC158639 Mus muscu
c 972 15.8 83.2 199277 12 AC165893 Bos tauru
c 973 15.8 83.2 199384 5 AP003531 Homo sapi
c 974 15.8 83.2 199426 11 CR847506 Zebrafish
c 975 15.8 83.2 200098 12 AC091161 Homo sapi
c 976 15.8 83.2 200192 12 AC109885 Rattus no
c 977 15.8 83.2 200202 6 AC112154 Mus muscu
c 978 15.8 83.2 200321 11 BX510646 Zebrafish
c 979 15.8 83.2 200439 6 AC160032 Mus muscu
c 980 15.8 83.2 200591 6 CT025604 Mouse DNA
c 981 15.8 83.2 200647 12 AL954633 Mus muscu
c 982 15.8 83.2 200796 12 AC120910 Rattus no
c 983 15.8 83.2 200985 6 AC121787 Mus muscu
c 984 15.8 83.2 201133 12 AL807235 Homo sapi
c 985 15.8 83.2 201466 12 AC147573 Homo sapi
c 986 15.8 83.2 202162 12 AC129627 Rattus no
c 987 15.8 83.2 202342 6 AL671867 Mouse DNA
c 988 15.8 83.2 203403 6 AC157789 Mus muscu
c 989 15.8 83.2 203589 6 AC084383 Mus muscu
c 990 15.8 83.2 203626 12 AC138730 Bongo PV9
c 991 15.8 83.2 203732 12 AC129845 Rattus no
c 992 15.8 83.2 204078 12 AC178149 Strongylo
c 993 15.8 83.2 204462 12 AC155619 Zea mays
c 994 15.8 83.2 204498 12 AC159547 Mus muscu
c 995 15.8 83.2 205357 12 AC152018 Colobus g
c 996 15.8 83.2 205900 6 AC103610 Mus muscu
c 997 15.8 83.2 206004 4 AP008246 Oryza sat
c 998 15.8 83.2 206096 12 AC181811 Strongylo
c 999 15.8 83.2 206125 6 AC153382 Mus muscu
c1000 15.8 83.2 206246 12 AC132970 Rattus no

```

## ALIGNMENTS

```

RESULT 1
BD233463/c 603 bp DNA linear PAT 17-JUL-2003
LOCUS
DEFINITION Human protein having hydrophobic domain and DNA encoding the same.
ACCESSION BD233463
VERSION BD233463.1 GI:33043233
KEYWORDS JP 2002519016-A/9.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 603)
Kato,S. and Kimura,T.
TITLE Human protein having hydrophobic domain and DNA encoding the same
JOURNAL Patent: JP 2002519016-A 9 02-JUL-2002;
SAGAMI CHEMICAL RESEARCH CENTER,PROTEGENE INC
COMMENT OS Homo sapiens (human)
PN JP 2002519016-A/9
PD 02-JUL-2002
PF 18-JUN-1999 JP 2000557267
PI SEISHI KATO,TOMOKO KIMURA
PC C12N15/09,C07K14/47,C12N1/15,C12N1/19,C12N5/10,C12N15/00,C12N5/ PC
00
CC Human protein having hydrophobic domain and DNA encoding the
FH same
FT Key Location/Qualifiers
1..603
/organism="Homo sapiens (human)"
/locus="BD233463"
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
FEATURES
source

```

```

ORIGIN
Query Match 100.0%; Score 19; DB 2; Length 603;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19
|||||
Db 591 CTTTACTTCATAGTCTTTG 573

RESULT 2
BD209699/c 1212 bp DNA linear PAT 17-JUL-2003
LOCUS
DEFINITION Compositions isolated from skin cells and methods for their use.
ACCESSION BD209699
VERSION BD209699.1 GI:33019469
KEYWORDS JP 2002512798-A/171.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 1212)
Strachan,L., Sleeman,M., Watson,J.D., Onrust,R., Kumble,A. and
Murison,J.G.
TITLE Compositions isolated from skin cells and methods for their use
JOURNAL Patent: JP 2002512798-A 171 08-MAY-2002;
GENESIS RESEARCH AND DEVELOPMENT CORP LTD
COMMENT OS Homo sapiens (human)
PN JP 2002512798-A/171
PD 08-MAY-2002
PF 29-APR-1999 JP 2000546009
PR 29-APR-1998 US 09/069726,09-NOV-1998 US 09/188930 PI
LORNA STRACHAN,MATTHEW SLEEMAN,JAMES DOUGLAS WATSON,RENE PI
ONRUST,
PI ANAND KUMBLE,JAMES GREG MURISON
PC C12N15/09,A61K38/00,A61P9/00,A61P17/00,A61P29/00,A61P31/18, PC
A61P35/00,
PC
C07K14/47,C12N1/15,C12N1/19,C12N1/21,C12N5/10,C12P21/02,C12N15/ PC
00, A61K37/02,C12N5/00
CC Compositions isolated from skin cells and methods for their
use.
FT Key Location/Qualifiers
1..1212
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
FEATURES
source
1..1212
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 100.0%; Score 19; DB 2; Length 1212;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19
|||||
Db 1031 CTTTACTTCATAGTCTTTG 1013

RESULT 3
AR341505/c 1212 bp DNA linear PAT 17-AUG-2003
LOCUS
DEFINITION Sequence 249 from patent US 6573095.
ACCESSION AR341505
VERSION AR341505.1 GI:33733640
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unknown.

```

```

Unclassified.
REFERENCE 1 (bases 1 to 1212)
AUTHORS Strachan,L.
TITLE Polynucleotides isolated from skin cells
JOURNAL Patent: US 6573095-A 249 03-JUN-2003;
Genesis Research & Development Corporation Limited; Parnell;
NZX;
FEATURES
source Location/Qualifiers
1..1212
/mol_type="genomic DNA"
ORIGIN
Query Match 100.0%; Score 19; DB 2; Length 1212;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 CTTTACTTCATAGTCTTTG 19
|||||
Db 1031 CTTTACTTCATAGTCTTTG 1013

RESULT 4
AX078375/c
LOCUS AX078375 1619 bp DNA linear PAT 22-FEB-2001
DEFINITION Sequence 43 from Patent WO0107612.
ACCESSION AX078375
VERSION AX078375.1 GI:13158044
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1
AUTHORS Au-Young,J., Bandman,O., Tang,Y.T., Yue,H., Azimzai,Y., Burford,N.,
Baughn,M.R., Lu,D.A., Hillman,J.L., Patterson,C. and Lal,P.
TITLE Receptors and associated proteins
JOURNAL Patent: WO 0107612-A 43 01-FEB-2001;
Incyte Genomics, Inc. (US)
FEATURES
source Location/Qualifiers
1..1619
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/notes="Incyte ID No: 2681738CBI".
ORIGIN
Query Match 100.0%; Score 19; DB 2; Length 1619;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 CTTTACTTCATAGTCTTTG 19
|||||
Db 1178 CTTTACTTCATAGTCTTTG 1160

RESULT 5
BD233473/c
LOCUS BD233473 1718 bp DNA linear PAT 17-JUL-2003
DEFINITION Human protein having hydrophobic domain and DNA encoding the same.
ACCESSION BD233473
VERSION BD233473.1 GI:33043243
KEYWORDS JP 2002519016-A/19.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 1718)
AUTHORS Kato,S. and Kimura,T.
TITLE Human protein having hydrophobic domain and DNA encoding the same
JOURNAL Patent: JP 2002519016-A 19 02-JUL-2002;

```

```

SAGAMI CHEMICAL RESEARCH CENTER, PROTEGENE INC
OS Homo sapiens (human)
PN JP 2002519016-A/19
PD 02-JUL-2002
PF 18-JUN-1999 JP 2000557267
PI SEISHI KATO,TOMOKO KIMURA
PC C12N15/09,C07K14/47,C12N1/15,C12N1/19,C12N5/10,C12N15/00,C12N5/ PC
00
CC Human protein having hydrophobic domain and DNA encoding the
CC same
FH Key 1..1718 Location/Qualifiers
FT source /organism="Homo sapiens (human)".
FEATURES
source Location/Qualifiers
1..1718
/mol_type="genomic DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 100.0%; Score 19; DB 2; Length 1718;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 CTTTACTTCATAGTCTTTG 19
|||||
Db 602 CTTTACTTCATAGTCTTTG 584

RESULT 6
AK172760/c
LOCUS AK172760 1826 bp mRNA linear PRI 07-MAY-2004
DEFINITION Homo sapiens cDNA FLJ3921 fis, clone COL02043, highly similar to
Homo sapiens retinoic acid induced 3 (RAI3).
ACCESSION AK172760
VERSION AK172760.1 GI:47077732
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1
AUTHORS Kawabata,A., Hikiiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,
Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1826)
AUTHORS Sugano,S. and Suzuki,Y.
TITLE Direct Submission
JOURNAL Submitted (22-APR-2004) Sumio Sugano, Institute of Medical Science,
University of Tokyo, Laboratory of Genome Structure, Human Genome
Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
(E-mail:filedna@mail.ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing;
Research Association for Biotechnology; cDNA library construction,
5'- & 3'-end one pass sequencing; Department of Virology and Human
Genome Center, Institute of Medical Science, University of Tokyo
(partly supported by Science and Technology Agency).
FEATURES
source Location/Qualifiers
1..1826
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="COL02043"
/tissue_type="colon"
/clone_lib="COL"
/note="cloning vector pME18SFL3"
ORIGIN

```

```

Query Match      100.0%; Score 19; DB 5; Length 1826;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCCTTTG 19
Db 679 CTTTACTTCATAGTCCTTTG 661

RESULT 7
CQ723177/c
LOCUS CQ723177 2290 bp DNA linear PAT 03-FEB-2004
DEFINITION Sequence 9111 from Patent WO02068579.
ACCESSION CQ723177
VERSION CQ723177.1 GI:42284034
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
REFERENCE 1
AUTHORS Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.
TITLE KITS, such as nucleic acid arrays, comprising a majority of
humanexons or transcripts, for detecting expression and other uses
thereof
JOURNAL Patent: WO 02068579-A 9111 06-SEP-2002;
PE Corporation (NY) (US)
FEATURES
source
1. .2290
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match      100.0%; Score 19; DB 2; Length 2290;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCCTTTG 19
Db 1158 CTTTACTTCATAGTCCTTTG 1140

RESULT 8
BC003665/c
LOCUS BC003665 2296 bp mRNA linear PRI 24-NOV-2004
DEFINITION Homo sapiens G protein-coupled receptor, family C, group 5, member
A, mRNA (cDNA clone MGC:923 IMAGE:2988011), complete cds.
ACCESSION BC003665
VERSION BC003665.2 GI:33872669
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
REFERENCE 1 (bases 1 to 2296)
AUTHORS Srausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Datchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Rana, S.S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalilus, D.E.,
Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 2296)
DIRECTOR MGC Project.
DIRECT SUBMISSION
Submitted (26-FEB-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgi.nci.nih.gov
On Aug 19, 2003 this sequence version replaced gi:13177795.
Contact: MGC help desk
Email: cgabps-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbio.org
contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahy, Erin Helton, Mark Kettman, Anuradha
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 3 Row: 1 Column: 1
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 12056470.
FEATURES
source
1. .2296
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:923 IMAGE:2988011"
/tissue_type="Colon, adenocarcinoma"
/clone_lib="NIH MGC 15"
/lab_host="DH10B-R"
/notes="Vector: pOTB7"
1. .2296
/genes="GPCR5A"
/notes="synonyms: RAIG1, GPCR5A, RAI3"
/db_xref="GeneID:9052"
/db_xref="MIM:604138"
95_..1168
/genes="GPCR5A"
/codon_start=1
/product="G protein-coupled receptor, family C, group 5,
member A"
/protein_id="AAH03665.1"
/db_xref="GI:13177796"
/db_xref="GeneID:9052"
/db_xref="MIM:604138"
/translation="MATTVPDGGCRNGLSKSYRLCDKAEAWGIVLETATAGVVTVA
FMLTPLIVCKQVDSNRKMLPTQFLFGLVGLFGLTFAPIIGLDGSGTPTREFLFG
ILFISCFCLAHAVSITKLVRGKPLSLVILGLAVGFSLVQDVIAIEVLVTMRT
NNVFWELASPRNEDFVLLLTLYVLFMLTFLMSSFTFCGSGTWKRGHAHLYTML
LSIAIWMATLLMLDFDRRWDITLSSALAANGVFLIAYVSPFWLLTKQNPM
YPVEDAFCKPQLVKYSYGVENRAYSQEITQGFETGDTLIAYPSTHFLQNPQPOKE
FSIPRAHAWSPYKDYEVKKEGS"

ORIGIN
Query Match      100.0%; Score 19; DB 5; Length 2296;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCCTTTG 19
Db 1153 CTTTACTTCATAGTCCTTTG 1135

```

```

RESULT 9
AF506289/c
LOCUS Homo sapiens orphan G protein-coupling receptor PEIG-1 mRNA, PRI 15-JUL-2002
DEFINITION complete cds.
ACCESSION AF506289
VERSION AF506289.1 GI:21779962
KEYWORDS Homo sapiens (human)
SOURCE
ORGANISM Homo sapiens (human)
REFERENCE 1 (bases 1 to 2297)
AUTHORS Cafarella,E.G., Gonzalez-Guerrico,A.M., Costanzo,R., Pivetta,O.H. and Santa-Coloma,T.A.
TITLE Identification by differential display of a mRNA specifically induced by 12-O-tetradecanoylphorbol-13-acetate (TPA) in T84 human colon carcinoma cells
JOURNAL Cell. Mol. Biol. 42 (5), 797-804 (1996)
PUBMED 8832110
REFERENCE 2 (bases 1 to 2297)
AUTHORS Cafarella,E.G., Gonzalez-Guerrico,A.M., Costanzo,R., Pivetta,O.H. and Santa-Coloma,T.A.
TITLE Direct Submission
JOURNAL Submitted (24-APR-2002) Laboratorio de Biologia Celular y Molecular, Instituto de Investigaciones Bioquimicas Fundacion Campomar, Patricia Argentinas 435, Buenos Aires 1405, Argentina
FEATURES
source
1. .2297
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="12"
/map="12p13-p12.3"
/sex="male"
/cell_line="T84"
/tissue_type="colorectal carcinoma"
/notes="derived from dbEST BE519991"
108. .1181
/notes="belongs to the type 3 G protein-coupling receptor family; characterized by a 7-transmembrane domain motif; phorbol ester induced protein-1"
/codon_start=1
/product="orphan G protein-coupling receptor PEIG-1"
/protein_id="AA077594.1"
/db_xref="GI:21779963"
/translation="MATTVPDGRNGLSKYRLCDKAEAGIVLETVATAGVVTVA
FMLTLPLVCKVQDSNRKMLPTQFLPLGLVGLFGLTFAPILGLDSTGPTFFLFG
ILFSICSLLAHAVALTKLVGRKPLSLVILGLAVGFLSLVQDVAIEIVLTNNRT
NNVFSLSAPRRNEDFVLLTYVFLMALTFMLSSFTFCGSGTGWKRHGAHLYLTM
LSIAIWVAWITLMLPDRDWDPTILSSALAANGWVFLAYVSPFEWLLTKQNPMD
YPVEDAFCKPOLVKKGVENRAYSQBEITQGETGDTLYAPYSTHFLQNQPPQKE
FSIPRAHAWSPYKDYEVKKEGS"
CDS
108. .1181
/notes="belongs to the type 3 G protein-coupling receptor family; characterized by a 7-transmembrane domain motif; phorbol ester induced protein-1"
/codon_start=1
/product="orphan G protein-coupling receptor PEIG-1"
/protein_id="AA077594.1"
/db_xref="GI:21779963"
/translation="MATTVPDGRNGLSKYRLCDKAEAGIVLETVATAGVVTVA
FMLTLPLVCKVQDSNRKMLPTQFLPLGLVGLFGLTFAPILGLDSTGPTFFLFG
ILFSICSLLAHAVALTKLVGRKPLSLVILGLAVGFLSLVQDVAIEIVLTNNRT
NNVFSLSAPRRNEDFVLLTYVFLMALTFMLSSFTFCGSGTGWKRHGAHLYLTM
LSIAIWVAWITLMLPDRDWDPTILSSALAANGWVFLAYVSPFEWLLTKQNPMD
YPVEDAFCKPOLVKKGVENRAYSQBEITQGETGDTLYAPYSTHFLQNQPPQKE
FSIPRAHAWSPYKDYEVKKEGS"
ORIGIN
Query Match 100.0%; Score 19; DB 5; Length 2297;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 CTTTACTTCATAGTCTTTG 19
| | | | | | | | | | | | | | | | | | | | | |
Db 1166 CTTTACTTCATAGTCTTTG 1148

RESULT 10
AX930411/c
LOCUS AX930411 2302 bp DNA linear PAT 22-DEC-2003
DEFINITION Sequence 2 from Patent WO03087832.
ACCESSION AX930411
VERSION AX930411.1 GI:40312209

```

```

KEYWORDS Homo sapiens (human)
SOURCE
ORGANISM Homo sapiens
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1
AUTHORS Terrett,J.A.
TITLE Diagnosis of carcinoma using raig1 polypeptides
JOURNAL Patent: WO 03087832-A 2 23-OCT-2003;
Oxford GlycoSciences (UK) Limited (GB)
FEATURES
source
1. .2302
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 100.0%; Score 19; DB 2; Length 2302;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 CTTTACTTCATAGTCTTTG 19
| | | | | | | | | | | | | | | | | | | | | |
Db 1158 CTTTACTTCATAGTCTTTG 1140

RESULT 11
AF095448/c
LOCUS AF095448 2302 bp mRNA linear PRI 29-DEC-1998
DEFINITION Homo sapiens putative G protein-coupled receptor (RAIG1) mRNA, complete cds.
ACCESSION AF095448
VERSION AF095448.1 GI:4063889
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 2302)
AUTHORS Cheng,Y. and Lotan,R.
TITLE Molecular cloning and characterization of a novel retinoic acid-inducible gene that encodes a putative G protein-coupled receptor
JOURNAL J. Biol. Chem. 273 (52), 35008-35015 (1998)
PUBMED 9857033
REFERENCE 2 (bases 1 to 2302)
AUTHORS Cheng,Y. and Lotan,R.
TITLE Direct Submission
JOURNAL Submitted (27-SEP-1998) Tumor Biology, The University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX 77030, USA
FEATURES
source
1. .2302
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="12"
/map="12p12.3-p13"
1. .2302
/genes="RAIG1"
/notes="induced by retinoic acid"
100. .1173
/genes="RAIG1"
/codon_start=1
/product="putative G protein-coupled receptor"
/protein_id="AAC98506.1"
/db_xref="GI:4063890"
/translation="MATTVPDGRNGLSKYRLCDKAEAGIVLETVATAGVVTVA
FMTLTLVCKVQDSNRKMLPTQFLPLGLVGLFGLTFAPILGLDSTGPTFFLFG
ILFSICSLLAHAVALTKLVGRKPLSLVILGLAVGFLSLVQDVAIEIVLTNNRT
NNVFSLSAPRRNEDFVLLTYVFLMALTFMLSSFTFCGSGTGWKRHGAHLYLTM

```

LSTAIWVAMITLMLDPDRRRDDTILSSALAANGVFLLAYVSPFWLLTKORNPMW  
YPVEDAFCKPQLVKSGYVENRAYSQEEITQGFETGDTLYAPYSTHFLQNPQPOKE  
FSIPRAHAWSPYKDYEVKKEGS"

ORIGIN

Query Match 100.0%; Score 19; DB 5; Length 2302;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCCTTTG 19

Db 1158 CTTTACTTCATAGTCCTTTG 1140

RESULT 12

BD156680/c 2446 bp DNA linear PAT 17-JAN-2003  
LOCUS  
DEFINITION Primer for synthesizing full-length cDNA and use thereof.

ACCESSION BD156680

VERSION BD156680.1 GI:27862438

KEYWORDS JP 2002191363-A/11523.

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.

1 (bases 1 to 2446)

Ota,T., Isoqai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,

Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.

Primer for synthesizing full-length cDNA and use thereof

Patent: JP 2002191363-A 11523 09-JUL-2002;

HELIX RESEARCH INSTITUTE

OS Homo sapiens (human)

PN JP 2002191363-A/11523

PD 09-JUL-2002

PF 28-JUL-2000 JP 2000280990

PI TOSHIO OTA,TAKAO ISOQAI,TETSUO NISHIKAWA,KOJI HAYASHI,KAORU

PI SAITO,

PI JUNICHI YAMAMOTO,SHIZUKO ISHII,TOMOYASU SUGIYAMA,AI WAKAMATSU,

PI KEIICHI NAGAI,TETSUJI OTSUKI

PC C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/PC

10,

PC C12P21/02,C12Q1/68/C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC

Primer for synthesizing full-length cDNA and use thereof FH Key

Location/Qualifiers (254). (1324).

FT CDS

Location/Qualifiers

FEATURES

source

1. .2446

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

ORIGIN

Query Match 100.0%; Score 19; DB 2; Length 2446;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCCTTTG 19

Db 1312 CTTTACTTCATAGTCCTTTG 1294

RESULT 13

AX877483/c

LOCUS AX877483 2446 bp DNA linear PAT 17-DEC-2003

DEFINITION Sequence 12388 from Patent EP1074617.

ACCESSION AX877483

VERSION AX877483.1 GI:40032219

KEYWORDS Homo sapiens (human)

SOURCE Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE

AUTHORS

1

Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.

Primer for synthesizing full-length cDNA and their use

Patent: EP 1074617-A 12388 07-FEB-2001;

Research Association for Biotechnology (JP)

Location/Qualifiers

1. .2446

/organism="Homo sapiens"

/mol\_type="unassigned DNA"

/db\_xref="taxon:9606"

254. .1327

/note="unnamed protein product"

/codon\_start=1

/protein\_id="CAE89746.1"

/db\_xref="GI:40032220"

/translation="MATTFDGCNRGLKSKYYRLCDKARAWGIVLETVATAGVVTVA

FMLTFLILVKVQDSNRRKMLPTQFLILGLVIGLTFATFIIGLDSGTPTFFFLPG

ILFSICFCLLAHVAVSLKLVGRKPLSLVILGLAVGSLVQDVIAIEIVILVMNRT

NVNVSLSAPRRNDFVLLITYVFLMALTFMLMSFTFCGSGTKWRGHGHIYLTML

LSTAIWVAMITLMLDPDRRRDDTILSSALAANGVFLLAYVSPFWLLTKORNPMW

YPVEDAFCKPQLVKSGYVENRAYSQEEITQGFETGDTLYAPYSTHFLQNPQPOKE

FSIPRAHAWSPYKDYEVKKEGS"

ORIGIN

Query Match 100.0%; Score 19; DB 2; Length 2446;

Best Local Similarity 100.0%; Pred. No. 36;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCCTTTG 19

Db 1312 CTTTACTTCATAGTCCTTTG 1294

RESULT 14

AK001761/c

LOCUS

DEFINITION Homo sapiens cDNA FLJ10899 fis, clone NT2RP5003506.

ACCESSION AK001761

VERSION AK001761.1 GI:7023229

KEYWORDS oligo capping; fis (full insert sequence).

SOURCE Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.

1

Ota,T., Suzuki,Y., Nishikawa,T., Otsuki,T., Sugiyama,T., Irie,R.,

Wakamatsu,A., Hayashi,K., Saito,H., Nagai,K., Kimura,K., Makita,H.,

Sekine,M., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,

Ishii,S., Yamamoto,J., Saito,K., Kawai,Y., Isono,Y., Nakamura,Y.,

Nagahari,K., Murakami,K., Yasuda,T., Iwayanagi,T., Wagatsuma,M.,

Shiratori,A., Sudo,H., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H.,

Sugawara,M., Takahashi,M., Kanda,K., Kamiyama,K., Katsuta,N., Sato,K.,

Kikkawa,B., Okura,Y., Abe,K., Tanai,H., Kimata,M., Watanabe,M.,

Tanikawa,M., Yamazaki,M., Ninomiya,K., Ishibashi,T., Yamashita,H.,

Murakawa,K., Fujimori,K., Tanai,H., Kimata,M., Watanabe,M.,

Hiraoka,S., Chiba,Y., Ishida,S., Ono,Y., Takiguchi,S., Watanabe,S.,

Yosida,M., Hotuta,T., Kusano,J., Kanehori,K., Takahashi-Fujii,A.,

Hara,H., Tanase,T.O., Nomura,Y., Togiya,S., Komai,F., Hara,R.,

Takeuchi,K., Arita,M., Imose,N., Musashino,K., Yuuki,H., Oshima,A.,

Sasaki,N., Aotsuka,S., Yoshikawa,Y., Matsunawa,H., Ichihiro,T.,

Shiohata,N., Sano,S., Moriya,S., Momiyama,H., Satoh,N., Takami,S.,

Terashima,Y., Suzuki,O., Nakagawa,S., Senoh,A., Mizoguchi,H.,

Goto,Y., Shimizu,F., Wakebe,H., Hishigaki,H., Watanabe,T.,

Sugiyama,A., Takemoto,M., Kawakami,B., Yamazaki,M., Watanabe,K.,

Kumagai,A., Itakura,S., Fukuzumi,Y., Fujimori,Y., Komiyama,M.,

Tashiro,H., Tanigami,A., Fujiwara,T., Ono,T., Yamada,K., Fujii,Y.,

Ozaki,K., Hirao,M., Ohmori,Y., Kawabata,A., Hikiji,T., Kobatake,N.,

Inagaki,H., Ikema,Y., Okamoto,S., Okitani,R., Kawakami,T.,

Noguchi,S., Itoh,T., Shigeta,K., Senba,T., Matsumura,K.,

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.

1

Ota,T., Isoqai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,

Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.

Primer for synthesizing full-length cDNA and their use

Patent: EP 1074617-A 12388 07-FEB-2001;

Research Association for Biotechnology (JP)

Location/Qualifiers

1. .2446

/organism="Homo sapiens"

/mol\_type="unassigned DNA"

/db\_xref="taxon:9606"

254. .1327

/note="unnamed protein product"

/codon\_start=1

/protein\_id="CAE89746.1"

/db\_xref="GI:40032220"

/translation="MATTFDGCNRGLKSKYYRLCDKARAWGIVLETVATAGVVTVA

FMLTFLILVKVQDSNRRKMLPTQFLILGLVIGLTFATFIIGLDSGTPTFFFLPG

ILFSICFCLLAHVAVSLKLVGRKPLSLVILGLAVGSLVQDVIAIEIVILVMNRT

NVNVSLSAPRRNDFVLLITYVFLMALTFMLMSFTFCGSGTKWRGHGHIYLTML

LSTAIWVAMITLMLDPDRRRDDTILSSALAANGVFLLAYVSPFWLLTKORNPMW

YPVEDAFCKPQLVKSGYVENRAYSQEEITQGFETGDTLYAPYSTHFLQNPQPOKE

FSIPRAHAWSPYKDYEVKKEGS"

Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togaishi, T., Oyamada, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J., Satoh, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K., Nagase, T., Nomura, N., Kikuchi, H., Masuho, Y., Yamashita, R., Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S. Complete sequencing and characterization of 21,243 full-length human cDNAs  
Nat. Genet. 36 (1), 40-45 (2004)

2  
Isogai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y., Nishikawa, T., Negai, K., Sugano, S., Ishibashi, T., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M., Hiraoka, S., Ishii, S., Kawai, Y., Saito, K., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagahara, K., Masuho, Y. and Kanehori, K.  
NEDO human cDNA sequencing project  
Unpublished  
3 (bases 1 to 2446)  
Isogai, T. and Otsuki, T.  
Direct Submission  
Submitted (16-FEB-2000) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana Kiearazu Chiba 292-0812, Japan (E-mail: flj-cdn@nifty.com, Tel: 81-438-52-3975, Fax: 81-438-52-3986)  
NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing and clone selection; Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.  
Location/Qualifiers  
1. .2446  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="NT2RP5003506"  
/cell\_line="NT2"  
/cell\_type="teratocarcinoma"  
/clone\_lib="NT2RP5"  
/note="Cloning vector: pME18SFL3  
mRNA from NT2 neuronal precursor cells after 2-weeks  
retinoic acid (RA) induction."  
254. .1327  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="BAA91890.1"  
/db\_xref="GI:7023230"  
/translation="MATVTPDGRNGLSKYYRLCDKAEAWGIVLETATAGVVTSA  
FMLTPIIVCKVQDSNRKMLPTQFLFLLGLVIGLTFGLTFAPFIIGLDGTGTRFLLFG  
ILFSICFSLAHAVSLTKLVGRKPLSLVILGLAVGFSLVQDVIAIEYIVLTMMRT  
NNVFSLSAPRNEFVLLITVYVLFMALTFELMSFTFCGSGTCWRHGAIYLTWL  
LSIAIWANITLLMLPDRDRDDTILSSALAANGWVFLAYVSPFWLLTYKQNPMD  
YVDEDAFCPOLVKVSKYGVENRAYSQBEITQGFETGDTLYAPYSTHFLQNPQPOKE  
FSIPRAHWSPYKDYEVKKEGS"

CDS  
Query Match 100.0%; Score 19; DB 5; Length 2446;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19  
|||||  
Db 1312 CTTTACTTCATAGTCTTTG 1294

RESULT 15  
LOCUS CQ981495 2456 bp DNA linear PAT 25-JAN-2005  
DEFINITION Sequence 350 from Patent EP1498424.  
ACCESSION CQ981495  
VERSION CQ981495.1 GI:58190785  
KEYWORDS  
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
1  
REFERENCE  
AUTHORS Rosenthal, A., Hermann, K., Heiden, E., Pilarsky, C., Bruemendorf, T., Staub, E., Roepcke, S., Mennerich, D., Kinnemann, H. and Li, X.  
TITLE Human nucleic acid sequences from lung tumours  
JOURNAL Patent: EP 1498424-A 350 19-JAN-2005;  
Hinzmann, Bernd (DE); Hermann, Klaus (DE); Heiden, Esmeralda (DE); Rosenthal, Andre (DE)  
FEATURES  
source location/Qualifiers  
1. .2456  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

ORIGIN  
Query Match 100.0%; Score 19; DB 2; Length 2456;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19  
|||||  
Db 1312 CTTTACTTCATAGTCTTTG 1294

RESULT 16  
LOCUS DD210040/c 2456 bp DNA linear PAT 19-JAN-2006  
DEFINITION Methods of Diagnosis of Cancer, Compositions and Methods of Screening for Modulators of Cancer.  
ACCESSION DD210040  
VERSION DD210040.1 GI:85654022  
KEYWORDS JP 2005518782-A/34.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
1 (bases 1 to 2456)  
Zlotnik, A., Mack, D.H., Agiz, N., Gish, K.C., Hebeji, P.A., Wilson, K.E. and Afar, D.  
TITLE Methods of Diagnosis of Cancer, Compositions and Methods of Screening for Modulators of Cancer  
JOURNAL Patent: JP 2005518782-A 34 30-JUN-2005;  
COMMENT PROTEIN DESIGN LABS INC  
OS Homo sapiens  
PN JP 2005518782-A/34  
PD 30-JUN-2005  
PF 17-SEP-2002 JP 2003529912  
PR 12-APR-2002 US 60/372246, 08-FEB-2002 US 60/355257, PR 08-FEB-2002 US 60/355145, 13-NOV-2001 US 60/350666, PR 08-SEP-2001 US 60/323887, 17-SEP-2001 US 60/323469 PI albert zlotnik, david h mack, natasha agiz, kurt c gish, peter a pi hebeji, pi keith e wilson, daniel afar  
CC  
FH Key Location/Qualifiers.  
source location/Qualifiers  
1. .2456  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

ORIGIN  
Query Match 100.0%; Score 19; DB 2; Length 2456;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19  
|||||  
Db 1312 CTTTACTTCATAGTCTTTG 1294

```

RESULT 17
AX549168/c
LOCUS
DEFINITION Sequence 453 from Patent WO02061087.
ACCESSION AX549168
VERSION AX549168.1 GI:25813894
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE
1
AUTHORS Burmer,G.C., Roush,C.L. and Brown,J.P.
TITLE Antigenic peptides, such as for G protein-coupled receptors
(GPCRs) antibodies thereto, and systems for identifying such
antigenic peptides
JOURNAL Patent: WO 02061087-A 453 08-AUG-2002;
Lifespan Biosciences, Inc. (US)
FEATURES
source
1. .2456
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 100.0%; Score 19; DB 2; Length 2456;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCCTTTG 19
|||||
Db 1312 CTTTACTTCATAGTCCTTTG 1294

RESULT 18
AK122672/c
LOCUS
DEFINITION Homo sapiens cDNA FLJ16117 fis, clone ASTRO2003632, highly similar
to Homo sapiens putative G protein-coupled receptor (RAIG1) mRNA.
ACCESSION AK122672
VERSION AK122672.1 GI:34527861
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE
1
AUTHORS Ota,T., Suzuki,Y., Nishikawa,T., Otsuki,T., Sugiyama,T., Irie,R.,
Wakamatsu,A., Hayashi,K., Sato,H., Nagai,K., Kimura,K., Makita,H.,
Sekine,M., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,
Ishii,S., Yamamoto,J., Saito,K., Kawai,Y., Isono,Y., Nakamura,Y.,
Nagahari,K., Murakami,K., Yasuda,T., Iwayanagi,T., Wagatsuma,M.,
Shiratori,K., Sudo,H., Hosiiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
Sugawara,M., Takahashi,M., Kanda,K., Yokoi,T., Furuya,T.,
Kikkawa,E., Omura,Y., Abe,K., Kamihara,K., Katsuta,N., Sato,K.,
Tanikawa,M., Yamazaki,M., Ninomiya,K., Ishibashi,T., Yamashita,H.,
Murakawa,K., Fujimori,K., Tanai,H., Kimata,M., Watanabe,M.,
Hiraoka,S., Chiba,Y., Ishida,S., Ono,Y., Takiguchi,S., Watanabe,S.,
Yosida,M., Hotata,T., Kusano,J., Kanehori,K., Kanehori,K., Komai,F., Hara,R.,
Hara,H., Tanase,T.O., Nomura,Y., Togiya,S., Komai,F., Hara,R.,
Takeuchi,K., Arita,M., Imose,N., Musashino,K., Yuuki,H., Oshima,A.,
Sasaki,N., Aotsuka,S., Yoshikawa,Y., Matsunawa,H., Ichihara,T.,
Shiohata,N., Sano,S., Moriya,S., Momiyama,H., Satoh,N., Takami,S.,
Terashima,Y., Suzuki,O., Nakagawa,S., Senoh,A., Mizoguchi,H.,
Goto,Y., Shimizu,F., Wakebe,H., Hishigaki,H., Watanabe,T.,
Sugiyama,A., Takemoto,M., Kawakami,B., Yamazaki,M., Watanabe,K.,
Kumagai,A., Itakura,S., Fukuzumi,Y., Fujimori,Y., Komiyama,M.,
Tashiro,H., Tanigami,A., Fujiwara,T., Ono,T., Yanada,K., Fujii,Y.,
Ozaki,K., Hirao,M., Ohmori,Y., Kawabata,A., Hikiji,T., Kobatake,N.,
Inagaki,H., Ikema,Y., Okamoto,S., Okitani,K., Kawakami,T.,

```

```

Noguchi,S., Itoh,T., Shigeta,K., Senba,T., Matsumura,K.,
Nakajima,Y., Mizuno,T., Morinaga,M., Sasaki,M., Togashi,T.,
Oyama,M., Hata,H., Watanabe,M., Komatsu,T., Mizushima-Sugano,J.,
Sato,T., Shirai,Y., Takahashi,Y., Nakagawa,K., Okumura,K.,
Nagase,T., Nomura,N., Kikuchi,H., Masuho,Y., Yamashita,R.,
Nakai,K., Yada,T., Nakamura,Y., Ohara,O., Isogai,T. and Sugano,S.
Complete sequencing and characterization of 21,243 full-length
human cDNAs
Nat. Genet. 36 (1), 40-45 (2004)
JOURNAL
PUBMED 14702039
REFERENCE
2
AUTHORS Tanigami,A., Fujiwara,T., Shibahara,T., Goto,Y., Hirao,M.,
Shimizu,F., Wakebe,H., Ono,T., Hishigaki,H., Watanabe,T., Ozaki,K.,
Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S.,
Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T.,
Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M.,
Kikuchi,H., Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K.,
Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y.,
Sugano,S., Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
NEDO human cDNA sequencing project
Unpublished
3 (bases 1 to 3057)
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/cell_type="normal astrocytes (NHA5732)"
/clone_lib="ASTRO2"
/notes="cloning vector: pME18SFL3
primary culture, normal astrocytes"
FEATURES
source
1. .3057
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/cell_type="normal astrocytes (NHA5732)"
/clone_lib="ASTRO2"
/notes="cloning vector: pME18SFL3
primary culture, normal astrocytes"
ORIGIN
Query Match 100.0%; Score 19; DB 5; Length 3057;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCCTTTG 19
|||||
Db 1949 CTTTACTTCATAGTCCTTTG 1931

RESULT 19
AX188348/c
LOCUS
DEFINITION Sequence 4043 from Patent WO0142467.
ACCESSION AX188348
VERSION AX188348.1 GI:15139821
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE
1
AUTHORS Schlegel,R., Deeds,J., Berger,A. and Zhao,X.
TITLE Genes, compositions, kits, and methods for identification,
assessment, prevention, and therapy of cervical cancer

```



JOURNAL Patent: WO 0142467-A 4043 14-JUN-2001;  
Millennium Predictive Medicine, Inc. (US)

FEATURES  
source 1..3371  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

ORIGIN

Query Match 100.0%; Score 19; DB 2; Length 3371;  
Best Local Similarity 100.0%; Pred. No. 37;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCTTTG 19  
|||||

Db 1225 CTTTACTTCATAGTCTTTG 1207

RESULT 20  
CQ894732/c  
LOCUS CQ894732 6730 bp DNA linear PAT 05-NOV-2004  
DEFINITION Sequence 42 from Patent EP1471075.  
ACCESSION CQ894732  
VERSION CQ894732.1 GI:55467481  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.

REFERENCE 1  
AUTHORS Rosenthal, A.D., Pilarsky, C., Dahl, E., Specht, T., Brueemendorf, T.,  
Lichtner, R., Staub, E., Rospcke, S. and Li, X.I.  
TITLE Human nucleic acid sequences expressed in pancreatic carcinomas  
JOURNAL Patent: EP 1471075-A 42 27-OCT-2004;  
Hinzmann, Bernd (DE); Rosenthal, Andre (DE); Pilarsky, Christian  
(DE); Dahl, Edgar (DE); Specht, Thomas (DE); Lichtner, Rosemarie  
(DE)

FEATURES  
source Location/Qualifiers  
1..6730  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

ORIGIN

Query Match 100.0%; Score 19; DB 2; Length 6730;  
Best Local Similarity 100.0%; Pred. No. 38;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCTTTG 19  
|||||

Db 1312 CTTTACTTCATAGTCTTTG 1294

RESULT 21  
AC007688  
LOCUS AC007688 161577 bp DNA linear PRI 30-AUG-2002  
DEFINITION Homo sapiens 12 BAC RPC111-392P7 (Roswell Park Cancer Institute  
Human BAC Library) complete sequence.  
ACCESSION AC007688  
VERSION AC007688.15 GI:5815499  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.

REFERENCE 1 (bases 1 to 161577)  
AUTHORS Muzny, D., Arenson, A.D., Bouck, J., Bunac, C., Chen, Z., Ding, Y.,  
Dugan, S., Durbin, J., Forcum, J., Garcia, C., Gorrell, J.H.,  
Gorrell, L.L., Hernandez, J., Issat, A., Jackson, L., Kneitz, S.,  
Kondejewski, N., Lau, S., Leal, B., Lee, E., Lichtarge, O., Liu, W.,  
Logan, O., Lu, J., Marondel, I., Martinez, C., Merscher, S., Miller, A.,

Montgomery, K., Oswal, G., Pampell, L.R., Parish, B.J., Perez, L.,  
Rashid, N.D., Rives, C., Scherer, S.E., Shen, H., Shim, C., Simon, M.,  
Vo, Q., Williamson, A., Worley, K.C., Xiang, A.M., Yang, R., Yu, W.,  
Zhou, X., Kuchera, R., Nelson, D. and Gibbs, R.A.

Direct Submission  
Unpublished  
2 (bases 1 to 161577)

Worley, K.C.

Direct Submission

Submitted (01-JUN-1999) Molecular and Human Genetics, Baylor  
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 161577)

Worley, K.C.

Direct Submission

Submitted (01-SEP-1999) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

4 (bases 1 to 161577)

Worley, K.C.

Direct Submission

Submitted (16-MAY-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

5 (bases 1 to 161577)

Worley, K.C.

Direct Submission

Submitted (30-AUG-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

On Sep 1, 1999 this sequence version replaced gi:5757565.

INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email

gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the  
entire insert of this clone. Overlapping regions of clones are only  
sequenced and submitted once, so the sequence for the remainder of  
the insert may be found in the record for the adjacent clones.  
Overlapping clones are noted at the beginning and end of the  
Features listing.

#### ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches  
of a local database that includes entries from dbSTS, GDB, and  
local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green,  
unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST  
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the  
EST and cDNA sequences. Genes demonstrate at least two exons  
flanked by consensus splice sites that maintained sequence  
continuity across the splice junctions. Sequences that are not  
identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum  
standard of double strand coverage with a minimum of 2 clones and 2  
reads with no ambiguities or 2 chemistries with a minimum of 2  
clones and 3 reads with no ambiguities. If the sequence quality for  
a region does not meet this standard, it will be indicated in the  
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality  
standards - estimated error rate less than 1 per 10,000 bases.  
Reports of lowest quality individual bases and measures of base  
quality are listed below. Description of the metrics can be found  
at URL:

<http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

#### QUALSTAT-REPORT-----

----- Summary Statistics -----  
Contig length: 161577  
Phrap values in estimate: 160751

Average error rate (BCM-Phrap estimate): 0.000163681  
Fraction of Phrap values less than 40 : 0.0376047  
Number of consensus changing edits: 30  
Number of N's in consensus : 0

```
----- Consensus changing edits -----
Position Original+Context Edited+Context
7033 acctcgccgt (n) ccgccccctt acctcgccgt (c) ccgccccctt
47567 aaaaaaaa (n) gaaataaat aaaaaaaa (a) gaaataaat
51135 aaagaagaa (n) aaagaagaa aaagaagaa (a) aaagaagaa
75582 aaaaaaaa (n) aanggatgtt aaaaaaaa (a) aanggatgtt
75585 aaaaaaaa (n) gaatgttcgc aaaaaaaa (a) gaatgttcgc
75667 ctaagacaga (n) taagatctta ctaagacaga (a) taagatctta
75752 ttttaaatag (g) gctttgcttt ttttaaatag (t) gctttgcttt
84017 ggaaggaag (n) aaggaagag ggaaggaag (g) aaggaagag
85227 ttttttttt (n) tgtttttttt ttttttttt (t) tgtttttttt
95227 ggaagtcag (n) atgcagtgag ggaagtcag (g) atgcagtgag
99681 aatctcttat (n) ccgaattcca aatctcttat (g) ccgaattcca
111307 actagtatac (n) atcntttttt actagtatac (t) atcntttttt
111378 gtatacnatc (n) tttttttttt gtatacnatc (c) tttttttttt
112621 aaaaaaaa (n) ccatcctaga aaaaaaaa (a) ccatcctaga
135812 atcacccctc (c) tttttttttt atcacccctc (t) tttttttttt
137207 ttcgagggac (n) cgcaccacg ttcgagggac (a) cgcaccacg
137218 cgcaccacg (n) ctgggttaagt cgcaccacg (c) ctgggttaagt
145113 gaaagtga (n) natgtagaat gaaagtga (a) natgtagaat
145114 caagtgaan (n) atgtagaata caagtgaan (g) atgtagaata
145232 ggcacgggtg (n) ntaacnccag ggcacgggtg (c) ntaacnccag
145233 gacacgggtg (n) taacnccag gacacgggtg (c) taacnccag
145238 ggtgntcac (n) ccagtaatcc ggtgntcac (g) ccagtaatcc
145321 gcaacatgg (t) nmccccatcc gcaacatgg (g) nmccccatcc
145322 caacatgg (n) nmccccatcc caacatgg (a) nmccccatcc
145323 acatgggtt (n) nmcccatctc acatgggtt (a) nmcccatctc
145324 gttcaccagg (n) gttgtggcgt gttcaccagg (t) gttgtggcgt
145363 gggcggtg (t) tgcagccctt gggcggtg (g) tgcagccctt
145377 tgcctacttg (n) ggggggggat tgcctacttg (g) ggggggggat
145505 aaaaaaaaa (n) gaaaaaaaaa aaaaaaaaa (a) gaaaaaaaaa
```

----- Distribution of Quality < 40 Bases -----

	5	10	15	20	25	30	35	40
1000	*	*	*	*	*	*	*	*
900	*	*	*	*	*	*	*	*
800	*	*	*	*	*	*	*	*
700	*	*	*	*	*	*	*	*
600	*	*	*	*	*	*	*	*
500	*	*	*	*	*	*	*	*
400	*	*	*	*	*	*	*	*
300	*	*	*	*	*	*	*	*
200	*	*	*	*	*	*	*	*
100	*	*	*	*	*	*	*	*
0	*	*	*	*	*	*	*	*

Phrap Value Range

Version: 1.01 qxf0.  
Location/Qualifiers  
1. 161577  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="12"  
/clone="RP11-392P7"  
complement (3. .65)  
/rpt\_family="MER4B"  
161. .471  
/rpt\_family="AluSx"  
complement (607. .702)  
/rpt\_family="MIR"

FEATURES  
source

repeat\_region  
repeat\_region  
repeat\_region

repeat\_region  
repeat\_region  
repeat\_region

STS

/standard\_name="WIAP-759-STS"  
/db\_xref="dbSTS:65358"

repeat\_region

repeat\_region

repeat\_region

repeat\_region

misc\_feature

/notes="Region: Unigene cluster similar to AA056332 and  
AI074576"

STS

/standard\_name="SHGC-44583"  
/db\_xref="dbSTS:48451"

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

Query Match 100.0%; Score 19; DB 5; Length 161577;

Best Local Similarity 100.0%; Pred. No. 48;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCCTTTG 19

Db 94870 CTTTACTTCATAGTCCTTTG 94888

RESULT 22

RP43035C13

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1

Watanabe,H., Fujiyama,A., Hattori,M., Taylor,T.D., Toyoda,A.,

Kuroki,Y., Noguchi,H., Benkhalil,A., Lehrach,H., Sudbrak,R.,

Kube,M., Taenzer,S., Galocz,P., Platzer,M., Scharfe,M.,

Nordsiek,G., Blocker,H., Hellmann,I., Khaitovich,P., Paabo,S.,

Reinhardt,R., Zheng,H.J., Zhang,X.L., Zhu,G.F., Wang,B.F., Fu,G.,

Ren,S.X., Zhao,G.P., Chen,Z., Lee,Y.S., Cheong,J.E., Choi,S.H.,

Wu,K.M., Liu,T.T., Hsiao,K.J., Tsai,S.F., Kim,C.G., Oota,S.,

Kitano,T., Kohara,Y., Saitou,N., Park,H.S., Wang,S.Y., Yaspo,M.L.

and Sakaki,Y.

The Chimpanzee Chromosome 22 Sequencing Consortium

DNA sequence and comparative analysis of chimpanzee chromosome 22

Nature 429 (6990), 382-388 (2004)

PUBMED 15164055

RP43035C13 157305 bp DNA linear PRI 03-FEB-2005  
Pan troglodytes chromosome 22 BAC RP43-035C13, complete sequence.  
AL954230  
AL954230.2 GI:38453651

HTG. Pan troglodytes (chimpanzee)

Pan troglodytes

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Pan.

```
REFERENCE
AUTHORS   2 (bases 1 to 157305)
TITLE     The Chimpanzee Chromosome 22 Sequencing Consortium.
JOURNAL   Chimpanzee chromosome 22 genomic sequence
REFERENCE Unpublished
AUTHORS   3 (bases 1 to 157305)
TITLE     Scharfe, M., Berg, C., Conrad, A., Hornischer, K., Loehnert, T.H.,
JOURNAL   Ludwig, M., Thies, S., Weber, K. and Bloecker, H.
AUTHORS   Direct Submission
TITLE     Submitted (25-JUN-2003) GBF, Dept. of Genome Analysis, Mascheroder
JOURNAL   Weg 1, D-38124 Braunschweig, Germany. E-mail: info.genome@gbf.de
COMMENT   On Nov 19, 2003 this sequence version replaced gi:37620130.
          The Chimpanzee Chromosome 22 Sequencing Consortium consists of :
          *Chinese National Human Genome Center at Shanghai,
            Shanghai, China
          *GBF, Dept. of Genome Analysis, Braunschweig, Germany; *Institute
            of Molecular Biotechnology, Jena, Germany; *RIKEN Genome Research
            Center, Daejeon, Korea;
          *Max-Planck Institute for Molecular Genetics, Berlin, Germany;
          *National Institute of Genetics, Mishima, Japan;
          *National Yang Ming University Genome Research Center, Taipei,
            Taiwan;
          *RIKEN Genomic Sciences Center, Yokohama, Japan.
          PROGRAMS AND PARAMETERS USED FOR ANNOTATION:
          +-----+
          + Analysis and annotation were performed with the automatic
          + 'first-pass' annotation and submission tool
          + 'AnnoMitter' (Hornischer & Bloecker).
          + Programs used by 'AnnoMitter':
          +-----+
          +-----+ Genome Center
          +-----+ Center: GBF, Braunschweig
          +-----+ Web site: http://genome.gbf.de/
          +-----+ Contact: info.genome@gbf.de
          +-----+ Project Information
          +-----+ Center project name:
          +-----+ Center clone name: RP43-035C13
          +-----+ Sequencing vector: Summary Statistics
          +-----+ Chemistry: Dye-terminator-amersham: ##% of reads
          +-----+ Assembly program: Dye-primer-amersham: ##% of reads
          +-----+ Consensus quality: 0 bases at least Q40
          +-----+ Consensus quality: 0 bases at least Q30
          +-----+ Consensus quality: 0 bases at least Q20
          +-----+ Estimated insert size: ##; agarose-gel estimation
          +-----+ Estimated insert size: 157305; sum-of-contigs estimation
          +-----+
          *
          FEATURES
          source
          Location/Qualifiers
            1..157305
              /organism="Pan troglodytes"
              /mol_type="genomic DNA"
              /db_xref="taxon:9598"
              /chromosome="22"
              /clone="RP43-035C13"
            misc_feature
              1..157305
                /note="assembly_fragment-clone_end:T7-vector_side:left
                assembly_fragment-clone_end:SP6-vector_side:right"
          ORIGIN
            Query Match          94.7%; Score 18; DB 5; Length 157305;
            Best Local Similarity 100.0%; Pred. No. 1.7e+02;
            Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
            Qy 1 CTTTACTTCATAGTCTTT 18
               |||||
            Db 84071 CTTTACTTCATAGTCTTT 84088
          RESULT 23
          AF064865
          LOCUS AF064865 159424 bp DNA linear PRI 02-JUN-1998
```

```
DEFINITION Homo sapiens chromosome 21q22.3 PAC 58D10, complete sequence.
ACCESSION AF064865
VERSION AF064865.1 GI:3171160
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
MAMMALIA: Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 159424)
AUTHORS Blechschmidt, K., Dagand, E., Hildmann, T., Nordsiek, G., Drescher, B.,
TITLE Weber, J., Schattevoy, R., Yaspo, M.-L. and Rosenthal, A.
JOURNAL Direct Submission
COMMENT Submitted (12-MAY-1998) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
FEATURES
source
  1..159424
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
    /chromosome="21"
    /map="21q22.3"
    /complement(1224..1523)
    /inference="non-experimental evidence, no additional
    details recorded"
    /rpt_family="AluY"
    1263..1386
    /inference="non-experimental evidence, no additional
    details recorded"
    /note="GC score = 8.60 (124bp); Region: GC content"
    1932..2047
    /inference="non-experimental evidence, no additional
    details recorded"
    /note="W2EF, score = 92%"
    3299..3348
    /inference="non-experimental evidence, no additional
    details recorded"
    /note="Xpound exon prediction, score = 72% (0%)"
    complement(3508..3614)
    /inference="non-experimental evidence, no additional
    details recorded"
    /note="GenScan, score = -0.41%, comment = Terminal_exon
    107 bp frame: 0 phase: 2"
    3843..4124
    /inference="non-experimental evidence, no additional
    details recorded"
    /note="GRAIL, score = 53.000%, comment = good shadow"
    complement(3913..4089)
    /inference="non-experimental evidence, no additional
    details recorded"
    /note="Xpound exon prediction, score = 94% (0%)"
    complement(3913..4066)
    /inference="non-experimental evidence, no additional
    details recorded"
    /note="W2EF, score = 72%"
    complement(3913..4062)
    /inference="non-experimental evidence, no additional
    details recorded"
    /note="GRAIL, score = 93.000%, comment = excellent"
    complement(3913..4018)
    /inference="non-experimental evidence, no additional
    details recorded"
    /note="GenScan, score = 16.83%, comment = Initial_exon 106
    bp frame: 1 phase: 1"
    complement(4787..4829)
    /inference="non-experimental evidence, no additional
    details recorded"
    /note="Xpound exon prediction, score = 73% (0%)"
    complement(7468..7853)
    /inference="non-experimental evidence, no additional
    details recorded"
    /note="GenScan, score = 35.77%, comment = Terminal_exon
    386 bp frame: 0 phase: 2"
```



Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homidae; Homo.

## REFERENCE

## AUTHORS

1 (bases 1 to 179400)

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Bouckgalter, B., Brown, A., Camarata, J., Campopiano, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Fero, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Karatas, A., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Seaman, S., Severy, P., Sougne, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

Submitted (05-JAN-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 179400)  
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Bouckgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Fero, S., Ferreira, P., FitzGerald, M., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J., Meneus, L., Mihova, T., Menga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Seaman, S., Severy, P., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

Submitted (14-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
4 (bases 1 to 179400)  
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Bouckgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Fero, S., Ferreira, P., FitzGerald, M., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J., Meneus, L., Mihova, T., Menga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,

Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (25-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jun 25, 2002 this sequence version replaced gi:21426184.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L11939  
Center clone name: 540\_B\_6  
-----

Only the first 179.4 kilobases of this clone are being submitted.  
The remainder overlaps accession number AC004460 [WashU clone].

## FEATURES

## Source

1. 1179400  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="15"  
/map="15"  
/clones="RP11-540B6"  
/clone\_lib="RPC1-11 Human Male BAC"  
complement (1..459)  
/rpt\_family="L1MC4a"  
141..145  
/notes"<30 qual SNGL region"  
complement (500..646)  
/rpt\_family="L1MC3"  
complement (917..1123)  
/rpt\_family="MER58A"  
complement (1403..1687)  
/rpt\_family="AluSx"  
1808..1820  
/notes"<30 qual SNGL region"  
1830..1834  
/notes"<30 qual SNGL region"  
1852..1872  
/notes"<30 qual SNGL region"  
2190..2356  
/rpt\_family="MIR"  
complement (2576..2758)  
/rpt\_family="L2"  
3842..3911  
/rpt\_family="L2"  
complement (4527..4846)  
/rpt\_family="AluJo"  
7737..7796  
/rpt\_family="L3"  
complement (8202..8334)  
/rpt\_family="L2"  
10511..10809  
/rpt\_family="L2"  
complement (12144..12355)  
/rpt\_family="MIR"  
12395..12437  
/rpt\_family="(T)Gn"  
complement (12712..12851)  
/rpt\_family="L2"  
13689..13717  
/rpt\_family="AT\_rich"

```

repeat_region 15823..15909
/rpt_family="MERSB"
repeat_region complement(16744..17080)
/rpt_family="L1M4"
repeat_region 17280..17331
/rpt_family="(TCTA)n"
repeat_region complement(17332..17622)
/rpt_family="AluI"
repeat_region 17767..17894
/rpt_family="MER3"
repeat_region 19432..19731
/rpt_family="AluSx"
repeat_region 19824..19953
/rpt_family="(GGA)n"
repeat_region 20195..20238
/rpt_family="AT rich"
repeat_region complement(20286..20650)
/rpt_family="L1MC5"
repeat_region complement(20816..21056)
/rpt_family="L1MC5"
repeat_region 23098..23253
/rpt_family="L2"
repeat_region complement(23254..23531)
/rpt_family="AluSx"
repeat_region 23532..23809
/rpt_family="L2"
repeat_region 24283..24529
/rpt_family="L2"
repeat_region 24621..24723
/rpt_family="U6"
repeat_region 25405..25702
/rpt_family="AluSg"
repeat_region 25946..26049
/rpt_family="L2"
repeat_region complement(27440..27819)
/rpt_family="L1MD2"
repeat_region 27883..28175
/rpt_family="AluSx"
repeat_region 28188..28303
/rpt_family="FLAM_C"
repeat_region 28314..28606
/rpt_family="AluJo"
repeat_region 28652..28759
/rpt_family="Triggers"
repeat_region 29174..29468
/rpt_family="AluSx"

Query Match 94.7% Score 18; DB 5; Length 179400;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCCTT 18
|||||
Db 115416 CTTTACTTCATAGTCCTT 115433

RESULT 25
PTB144012
LOCUS PTB144012 180638 bp DNA linear PRI 03-FEB-2005
DEFINITION Pan troglodytes chromosome 22 BAC PTB-144012, complete sequence.
ACCESSION AL954231
VERSION AL954231.3 GI:38153749
KEYWORDS HTG.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Pan.
REFERENCE 1
Watanabe,H., Fujiyama,A., Hattori,M., Taylor,T.D., Toyoda,A.,
Kuroki,Y., Noguchi,H., Benkhalil,A., Leirach,H., Sudbrak,R.,
Kube,M., Taenzler,S., Galoczky,P., Platzner,M., Scharfe,M.,
Nordsiek,G., Blocker,H., Hellmann,I., Khaitovich,P., Paabo,S.,
Reinhardt,R., Zheng,H.J., Zhang,X.L., Zhu,G.F., Wang,B.F., Fu,G.,
Ren,S.X., Zhao,G.P., Chen,Z., Lee,Y.S., Cheong,J.B., Choi,S.H.,
Wu,K.M., Liu,T.T., Hsiao,K.J., Tsai,S.F., Kim,C.G., Oota,S.,
Kitano,T., Kohara,Y., Saitou,N., Park,H.S., Wang,S.Y., Yaspo,M.L.
and Sakaki,Y.
The Chimpanzee Chromosome 22 Sequencing Consortium
DNA sequence and comparative analysis of chimpanzee chromosome 22
Nature 429 (6990), 382-388 (2004)
CONSTRM
TITLE The Chimpanzee Chromosome 22 Sequencing Consortium.
JOURNAL Unpublished
PUBMED 15164055
REFERENCE 2 (bases 1 to 180638)
AUTHORS The Chimpanzee Chromosome 22 Sequencing Consortium.
TITLE Chimpanzee chromosome 22 genomic sequence
JOURNAL Unpublished
AUTHORS 3 (bases 1 to 180638)
Schäfer,M., Berg,C., Conrad,A., Hornischer,K., Loehnert,T.H.,
Ludwig,M., Thies,S., Weber,K. and Bloeker,H.
Direct Submission
Submitted (25-JUN-2003) GBF, Dept. of Genome Analysis, Mascheroder
Weg 1, D-38124 Braunschweig, Germany, E-mail: info.genome@gbf.de
On Nov 3, 2003 this sequence version replaced gi:37620132.
The Chimpanzee Chromosome 22 Sequencing Consortium consists of :
*Chinese National Human Genome Center at Shanghai,
Shanghai, China
*GBF, Dept. of Genome Analysis, Braunschweig, Germany; *Institute
of Molecular Biotechnology, Jena, Germany; *KRIIBB Genome Research
Center, Daejeon, Korea;
*Max-Planck Institute for Molecular Genetics, Berlin, Germany;
*National Institute of Genetics, Mishima, Japan;
*National Yang Ming University Genome Research Center, Taipei,
Taiwan;
*RIKEN Genomic Sciences Center, Yokohama, Japan.
PROGRAMS AND PARAMETERS USED FOR ANNOTATION:
*****
+ Analysis and annotation were performed with the automatic
+ 'first-pass' annotation and submission tool
+ 'AnnoMitter' (Hornischer & Bloeker).
+ Programs used by 'AnnoMitter':
*****
----- Genome Center
Center: GBF, Braunschweig
Center code: GBF
Web site: http://genome.gbf.de/
Contact: info.genome@gbf.de
----- Project Information
Center project name:
Center clone name: PTB-144012
----- Summary Statistics
Sequencing vector: ##;
Chemistry: Dye-terminator-amersham: ## of reads
Assembly: Dye-primer-amersham: ## of reads
Assembly program: ##
Consensus quality: 0 bases at least Q40
Consensus quality: 0 bases at least Q30
Consensus quality: 0 bases at least Q20
Estimated insert size: ##; agarose-fp estimation
Estimated insert size: 180638; sum-of-contigs estimation
*****
*
FEATURES
source Location/Qualifiers
1..180638
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/chromosome="22"
/clone="PTB-144012"
1..180638
misc_feature /note="assembly_fragment-clone_end:T7-vector_side:left
assembly_fragment-clone_end:SP6-vector_side:right"
ORIGIN
Query Match 94.7% Score 18; DB 5; Length 180638;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 1 CTTTACTTCATAGTCTTT 18
Db 42303 CTTTACTTCATAGTCTTT 42320

RESULT 26
AC073220 187647 bp DNA linear HTG 29-JUN-2000
LOCUS Homo sapiens chromosome 15 clone RP11-540B6, *** SEQUENCING IN
DEFINITION PROGRESS ***, 10 unordered pieces.
ACCESSION AC073220
VERSION AC073220.3 GI:8810532
KEYWORDS HTG: HTGS PHASE1.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (10-JUN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Jun 29, 2000 this sequence version replaced gi:8516745.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H NH0540B06
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-primer ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 181701 bases at least Q40
Consensus quality: 184075 bases at least Q30
Consensus quality: 185281 bases at least Q20
Insert size: 221000; agarose-fp
Insert size: 186747; sum-of-contigs
Quality coverage: 5.03 in Q20 bases; agarose-fp
Quality coverage: 5.97 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 2621: contig of 2621 bp in length
* 2622 2721: gap of unknown length
* 5795: contig of 3074 bp in length
* 5796 5895: gap of unknown length
* 13340: contig of 7445 bp in length
* 13341 13440: gap of unknown length
* 13441 22039: contig of 8599 bp in length
* 22040 22139: gap of unknown length
* 33967: contig of 11828 bp in length
* 33968 46186: gap of unknown length
* 46186 46285: gap of unknown length
* 46285 64460: gap of unknown length
* 64460 64559: gap of unknown length

* 64560 86907: contig of 22348 bp in length
* 86908 87007: gap of unknown length
* 87008 107166: contig of 20159 bp in length
* 107167 187647: contig of 80381 bp in length.
Location/Qualifiers
1. 187647
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="15"
/clone="RP11-540B6"
1. 2621
/note="assembly_name:Contig7"
2622..2721
/estimated_length=unknown
2722..5795
/note="assembly_name:Contig8"
clone end:SP6
vector_side:right"
5796..5895
/estimated_length=unknown
5896..13340
/note="assembly_name:Contig9"
clone end:T7
vector_side:left"
13341..13440
/estimated_length=unknown
13441..22039
/note="assembly_name:Contig10"
22040..22139
/estimated_length=unknown
22140..33967
/note="assembly_name:Contig11"
33968..34067
/estimated_length=unknown
34068..46185
/note="assembly_name:Contig12"
46186..46285
/estimated_length=unknown
46286..64459
/note="assembly_name:Contig13"
64460..64559
/estimated_length=unknown
64560..86307
/note="assembly_name:Contig14"
86308..87007
/estimated_length=unknown
87008..107166
/note="assembly_name:Contig15"
107167..107266
/estimated_length=unknown
107267..187647
/note="assembly_name:Contig16"

ORIGIN
Query Match 94.7%; Score 18; DB 12; Length 187647;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 CTTTACTTCATAGTCTTT 18
Db 54480 CTTTACTTCATAGTCTTT 54497

RESULT 27
RP43039B12 201761 bp DNA linear PRI 03-FEB-2005
LOCUS Pan troglodytes chromosome 22 BAC RP43-039B12, complete sequence.
DEFINITION Pan troglodytes chromosome 22 BAC RP43-039B12, complete sequence.
ACCESSION AL954232
VERSION AL954232.2 GI:37620133
KEYWORDS HTG.
SOURCE Pan troglodytes (chimpanzee)

```

```
ORGANISM  Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Pan.

REFERENCE 1
AUTHORS  Watanabe,H., Fujiyama,A., Hattori,M., Taylor,T.D., Toyoda,A.,
Kuroki,Y., Noguchi,H., Benkhalil,A., Lehrach,H., Sudbrak,R.,
Kube,M., Taenzer,S., Galoczzy,P., Platzer,M., Scharfe,M.,
Nordsiek,G., Blocker,H., Hellmann,I., Khaitovich,P., Paabo,S.,
Reinhardt,R., Zheng,H.J., Zhang,X.L., Zhu,G.F., Wang,B.F., Fu,G.,
Ren,S.X., Zhao,G.P., Chen,Z., Lee,Y.S., Cheong,J.E., Choi,S.H.,
Wu,K.M., Liu,T.T., Hsiao,K.J., Tsai,S.F., Kim,C.G., Ooka,S.,
Kitano,T., Kohara,Y., Saitou,N., Park,H.S., Wang,S.Y., Yaspo,M.L.
and Sakaki,Y.
The Chimpanzee Chromosome 22 Sequencing Consortium
The Chimpanzee sequence and comparative analysis of chimpanzee chromosome 22
NATURE 429 (6990), 382-388 (2004)
15164055
2 (bases 1 to 201761)
The Chimpanzee Chromosome 22 Sequencing Consortium.
Unpublished
3 (bases 1 to 201761)
Ludwig,M., Berg,C., Conrad,A., Hornischer,K., Loehnert,T.H.,
Scharfe,M., Thies,S., Weber,K. and Bloeker,H.
Direct Submission
Submitted (25-JUN-2003) GBF, Dept. of Genome Analysis, Mascheroder
Weg 1, D-38124 Braunschweig, Germany, E-mail: info.genome@gbf.de
The Chimpanzee Chromosome 22 Sequencing Consortium consists of :
*Chinese National Human Genome Center at Shanghai,
Shanghai, China
*GBF, Dept. of Genome Analysis, Braunschweig, Germany; *Institute
of Molecular Biotechnology, Jena, Germany; *KRIIBB Genome Research
Center, Daejeon, Korea;
*Max-Planck-Institute for Molecular Genetics, Berlin, Germany;
*National Institute of Genetics, Mishima, Japan;
*National Yang Ming University Genome Research Center, Taipei,
Taiwan;
*RIKEN Genomic Sciences Center, Yokohama, Japan.
PROGRAMS AND PARAMETERS USED FOR ANNOTATION:
+ Analysis and annotation were performed with the automatic
+ 'first-pass' annotation and submission tool
+ 'AnnoMitter' (Hornischer & Bloeker).
+ Programs used by 'AnnoMitter':
+
+----- Genome Center
Center: GBF, Braunschweig
Center code: GBF
Web site: http://genome.gbf.de/
Contact: info.genome@gbf.de
----- Project Information
Center project name:
Center clone name: RP43-039B12
----- Summary Statistics
Sequencing vector: ###;
Chemistry: Dye-terminator-amersham: ## of reads
Chemistry: Dye-primer-amersham: ## of reads
Assembly program: #
Consensus quality: 0 bases at least Q40
Consensus quality: 0 bases at least Q30
Consensus quality: 0 bases at least Q20
Estimated insert size: ##; agarose-fp estimation
Estimated insert size: 201761; sum-of-contigs estimation
+-----
* Location/Qualifiers
1. .201761
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/chromosome="22"
/clone="RP43-039B12"

misc_feature 1. .201761
/notes="assembly_fragment-clone_end:T7-vector_side:left
assembly_fragment-clone_end:SP6-vector_side:right"

ORIGIN
Query Match 94.7%; Score 18; DB 5; Length 201761;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CTTTACTTCATAGCTCTTT 18
|||||
Db 50509 CTTTACTTCATAGCTCTT 50526

RESULT 28
AC099102/c
AC099102/c
LOCUS Rattus norvegicus clone CH230-183A24, *** SEQUENCING IN PROGRESS
DEFINITION ***; 2 unordered pieces.
AC099102 206082 bp DNA linear HTG 10-MAY-2003
AC099102.6 GI:30522538
HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Muridae; Murinae; Rattus.
1 (bases 1 to 206082)
REFERENCE 1
AUTHORS Muzny,D., Marle,M., Metzker,M., Lee,A., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Bliswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buha,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Caesar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Delgado,O., Denson,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Ding,Y., Dinh,H., Divya,K.,
Davila,M.L., Davis,C., Davy-carroll,L., De Anda,C., Falls,T., Fan,G.,
Egan,A., Escoto,M., Eugene,C., Evans,C.A., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Georgiev,G., Geier,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Loulseghe,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmood,M., Malloy,K., Mangum,A.,
Mangum,B., Napua,P., Martin,K., Martin,R., Martinez,E.,
Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokemele,O., Okunolu,G., Olarnpunagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
Popper,F., Poindester,A., Popovic,D., Primus,E., Pu.L.L.,
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Snajds,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorrelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villalana,D., Waldron,L., Walker,B., Wang,J.,
Wang,O., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wlarczyk,R., Wooden,H., Worley,K.,
```



Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

Direct Submission  
Unpublished  
2 (bases 1 to 206082)  
Worley, K. C.

Direct Submission  
Submitted (09-NOV-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 206082)

Reference  
Rat Genome Sequencing Consortium.  
Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On May 10, 2003 this sequence version replaced gi:23264229. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: http://www.hgsc.bcm.tmc.edu/  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GGPI  
Center clone name: CH230-183A24  
----- Summary Statistics  
Assembly program: Atlas 3.0;  
Consensus quality: 193499 bases at least Q40  
Consensus quality: 195436 bases at least Q30  
Consensus quality: 196841 bases at least Q20  
Estimated insert size: 201950; sum-of-contigs estimation  
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
(see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 2 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 204625: contig of 204625 bp in length  
\* 204626 204725: gap of unknown length  
\* 204726 206082: contig of 1357 bp in length.

FEATURES  
source  
Location/Qualifiers  
1..206082  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-183A24"

misc\_feature  
1..1533  
/note="wgs\_end\_extension  
clone\_end:T7"  
2990..4234  
/note="wgs\_end\_extension  
clone\_end:T7"  
5188..5661

-----  
/note="clone\_boundary  
clone\_end:T7  
site:EcoRI  
end\_sequence:BH339247"  
complement(203280..203920)  
/note="clone\_boundary  
clone\_end:Sp6  
site:EcoRI  
end\_sequence:BH339260"  
204626..204725  
/estimated\_length=unknown

misc\_feature  
gap  
204626..204725  
/estimated\_length=unknown

ORIGIN  
Query Match 94.7%; Score 18; DB 12; Length 206082;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 2 TTTACTTCATAGTCTTTG 19  
|||||  
Db 172840 TTTACTTCATAGTCTTTG 172823

RESULT 29  
CT027583/c  
LOCUS  
DEFINITION  
Danio rerio chromosome 2 clone DKEY-23H1, WORKING DRAFT SEQUENCE, 7  
unordered pieces.  
CT027583  
VERSION  
CT027583.4 GI:76061021  
KEYWORDS  
HTG; HTGS PHASE1; HTGS DRAFT; HTGS\_FULLTOP.  
SOURCE  
Danio rerio (zebrafish)  
ORGANISM  
Danio rerio  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
Cypriniformes; Cyprinidae; Danio.  
1 (bases 1 to 249140)  
McLaren, S.  
Direct Submission  
Submitted (02-FEB-2006) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
zfish-help@sanger.ac.uk Clone requests:  
http://www.sanger.ac.uk/Projects/D\_rerio/faqs.shtml#dataeight  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: zfish-help@sanger.ac.uk  
----- Project Information  
Center project name: zK23H1  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Chemistry: Dye-terminator; 100% of reads  
Consensus quality: 246887 bases at least Q40  
Consensus quality: 247347 bases at least Q30  
Consensus quality: 247722 bases at least Q20  
Insert size: 248540; sum-of-contigs  
Insert size: 250444; 3.5% error; agarose-fp  
Quality coverage: 7.80x in Q20 bases; sum-of-contigs Quality  
coverage: 7.78x in Q20 bases; agarose-fp

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 7 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved

1 63859: contig of 63859 bp in length  
\* 63860 63959: gap of 100 bp  
\* 63960 68546: contig of 4587 bp in length  
\* 68547 68646: gap of 100 bp  
\* 68647 100429: contig of 31783 bp in length

```

* 100430 100528: gap of 100 bp
* 100530 107413: contig of 6884 bp in length
* 107414 107513: gap of 100 bp
* 107514 191566: contig of 84053 bp in length
* 191567 191666: gap of 100 bp
* 191667 224282: contig of 32616 bp in length
* 224283 224382: gap of 100 bp
* 224383 249140: contig of 24758 bp in length.
FEATURES
    source
        1. .249140
            /organism="Danio rerio"
            /mol_type="genomic DNA"
            /db_xref="RZPD:HUKB735H0123Q"
            /db_xref="taxon:7955"
            /chromosome="2"
            /clone="DKEY-23H1"
            /clone_lib="DanioKey"
        1. .63859
            /notes="assembly_fragment:01360"
            /fragment_chain:1"
            /notes="assembly_fragment:00017"
            /fragment_chain:1"
            /notes="assembly_fragment:00932"
            /fragment_chain:1"
            /notes="assembly_fragment:00080"
            /fragment_chain:1"
            /notes="assembly_fragment:02187"
            /fragment_chain:2"
            /notes="assembly_fragment:00549"
            /fragment_chain:2"
            /notes="assembly_fragment:00173.0"
ORIGIN
    Query Match          94.7%; Score 18; DB 12; Length 249140;
    Best Local Similarity 100.0%; Pred. No. 1.7e+02;
    Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 CTTTACTTCATAGTCCTTT 18
    |||||
Db 176396 CTTTACTTCATAGTCCTTT 176379
RESULT 30
HS21C081
LOCUS HS21C081 340000 bp DNA linear PRI 16-APR-2005
DEFINITION Homo sapiens chromosome 21 segment HS21C081.
ACCESSION AL163281 AP001736 BA000005
VERSION AL163281.2 GI:7717373
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
    Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
    Hominidae; Homo.
REFERENCE 1 (bases 1 to 340000)
AUTHORS Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
    Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,
    Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
    Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
    Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Schudy,A.,
    Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K.,
    Aakawa,S., Shintani,A., Sasaki,T., Nagamine,K., Mitsuyma,S.,
    Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordstiek,G.,
    Hornischer,K., Brandt,P., Scharfe,M., Schoen,O., Desario,A.,
    Reichelt,J., Kauer,G., Bloeker,H., Ramser,J., Beck,A., Klages,S.,
    Hennig,S., Riesselmann,L., Dagand,E., Wehrmeyer,S., Borzym,K.,
    Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and

```

```

Vaspo,M.I.
Direct Submission
Submitted (05-MAY-2000) The Chromosome 21 Mapping and Sequencing
Consortium: * RIKEN Genomic Sciences Center, Human Genome Research
Group * Institute of Molecular Biotechnology, Genome Analysis *
Keio University School of Medicine, Dept. of Molecular Biology *
GBF, Dept. of Genome Analysis * Max-Planck Institute for Molecular
Genetics (addresses see below)
The Chromosome 21 Mapping and Sequencing Consortium consists of
* RIKEN Genomic Sciences Center, Human Genome Research Group, *
Sagamihara 228-8555, Japan,
* e.mail: sakaki@gsc.riken.go.jp
* URL: http://hgp.gsc.riken.go.jp/
and
* Institute of Molecular Biotechnology, Genome Analysis, *
Beutenbergstrasse 11, D-07745 Jena, Germany,
* e.mail: gscj-submit@genome.imb-jena.de
* URL: http://genome.imb-jena.de/
and
* Keio University School of Medicine, Dept. of Molecular Biology, *
Tokyo 160-8582, Japan,
* e.mail: shimizu@dm-b-med.keio.ac.jp
* URL: http://adenine.dmb.med.keio.ac.jp/
and
* GBF, Dept. of Genome Analysis,
* Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e.mail:
info.genome@gbf.de
* URL: http://genome.gbf.de/
and
* Max-Planck Institute for Molecular Genetics,
* Ihnestrasse 73, D-14195 Berlin, Germany,
* e.mail: info-chr21@molgen.mpg.de
* URL: http://chr21.rz-berlin.mpg.de/.
FEATURES
    source
        1. .340000
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="21"
            /map="21q22.3"
            <1. .1000
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="21"
            /map="21q22.3"
            /clone="P206A10, 5' partial"
            /clone_lib="RPC11,3-5 PAC library"
            /note="Accession No. AF121782"
            <1. .147816
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="21"
            /map="21q22.3"
            /clone="BAC-291B3, 5' partial"
            /clone_lib="BAC library"
            /note="Accession No. AF064857"
            134734. .167600
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="21"
            /map="21q22.3"
            /clone="O7G4"
            /note="Accession No. AF045449"
            154163. .300059
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="21"
            /map="21q22.3"

```

```

/clones="P31P10"
/clone_lib="RPC11,3-5 PAC library"
/note="Accession No. AF064862"
235367. .340000
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone="P58D10, 3' partial"
/clone_lib="RPC11,3-5 PAC library"
/note="Accession No. AF064865"
2494. .2520
/note="(CAAAA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
2620. .3399
/note="L1MC4"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
complement(5949. .6158)
/note="AluJb"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(6294. .6451)
/note="L1MB5"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
7710. .7717
/note="Not1 site"
7942. .8206
/note="AluJo"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(8367. .8439)
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
8568. .8882
/note="AluSq"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(11216. .11244)
/note="MIR"
/rpt_family="SINE/MIR"
/rpt_type=DISPERSED
complement(11245. .11259)
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
11362. .11331
/note="MER16"
/rpt_family="DNA/Mariner"
/rpt_type=DISPERSED
complement(11341. .11465)
/note="MIR"
/rpt_family="SINE/MIR"
/rpt_type=DISPERSED
13420. .13527
/note="(TG)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
13716. .14123
/note="WL1D"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
16010. .16082
/note="MER67A"
/rpt_family="LTR/MER4-group"
/rpt_type=DISPERSED
16083. .16589
/note="MER67B"
/rpt_family="LTR/MER4-group"

```

```

repeat_region
16623. .16789
/note="MER67A"
/rpt_family="LTR/MER4-group"
/rpt_type=DISPERSED
17023. .17083
/note="AluSg"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(17728. .18041)
/note="AluJb"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(19405. .19488)
/note="Alu"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(19506. .20061)
/note="L1MEC"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
complement(20236. .20397)
/note="HERVL"
/rpt_family="LTR/Retroviral"
/rpt_type=DISPERSED
complement(20500. .20848)
/note="HERVL"
/rpt_family="LTR/Retroviral"
/rpt_type=DISPERSED
complement(20944. .21212)
/note="HERV16"

Query Match          94.7%; Score 18; DB 5; Length 340000;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTT 18
    |||||
Db 312022 CTTTACTTCATAGTCTTT 312039

RESULT 31
DQ271038/c
LOCUS      DQ271038          762 bp    DNA    linear    MAM 19-JAN-2006
DEFINITION Mustela vison clone VIIISN11 microsatellite sequence.
ACCESSION  DQ271038
VERSION    DQ271038.1 GI:82797063
KEYWORDS
SOURCE     Mustela vison (American mink)
ORGANISM   Mustela vison
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia;
            Mustelidae; Mustelinae; Mustela.
REFERENCE  1 (bases 1 to 762)
AUTHORS   Anistoraei,R. and Christensen,K.
TITLE     Isolation and characterization of 79 microsatellite makers for the
JOURNAL   American mink (Mustela vison)
REFERENCE  2 (bases 1 to 762)
AUTHORS   Anistoraei,R. and Christensen,K.
TITLE     Direct Submission
JOURNAL   Submitted (31-OCT-2005) Division of Animal Genetics, Department of
            Animal and Veterinary Basic Sciences, The Royal Veterinary and
            Agricultural University, Groennegaardsvej 3, Frederiksberg C 1870,
            Denmark
FEATURES
            Location/Qualifiers
            1..762
                organism="Mustela vison"
                mol_type="genomic DNA"
                db_xref="taxon:9667"
                clone="VIIISN11"
            1..762
                note="microsatellite"
repeat_region

```

```

ORIGIN
Query Match          91.6%; Score 17.4; DB 14; Length 762;
Best Local Similarity 94.7%; Pred. No. 2.4e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCCTTTG 19
|||||
Db 600 CTTTACTTCATAGTCCTTTG 582

RESULT 32
SCYGR131W          1744 bp DNA linear PLN 11-AUG-1997
LOCUS              S.cerevisiae chromosome VII reading frame ORF YGR131w.
DEFINITION         S.cerevisiae chromosome VII reading frame ORF YGR131w.
ACCESSION          272916 Y13135
VERSION            272916.1 GI:1323216
KEYWORDS
SOURCE
ORGANISM
Saccharomyces cerevisiae (baker's yeast)
Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE
AUTHORS            van Dyck,L., Skala,J., de Wergifosse,P., Purnelle,B., Talla,E.,
Nawrocki,A., Del Bino,S. and Goffeau,A.
JOURNAL            Unpublished
REFERENCE          2 (bases 1 to 1744)
MIPS
Direct Submission
Submitted (14-MAY-1996) Data collected by MIPS on behalf of the
European yeast chromosome VII sequencing project. MIPS at the
Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152
Martinsried, FRG; E-mail: Mewes@mips.embnnet.org
LOCATION/Qualifiers
1. .1744
/organism="Saccharomyces cerevisiae"
/mol_type="genomic DNA"
/db_xref="taxon:4932"
/chromosome="VII"
complement(744..754)
/notes="ARS-consensus"
misc_feature
882..1406
/notes="unnamed protein product; ORF YGR131w"
codon_start=1
/protein_id="CAA97144.1"
/db_xref="GI:1323217"
/db_xref="GOA:P53279"
/db_xref="SGD:S0003363"
/translation="MLFNTNRQDDPPTOLINOSANTONQQAHQESPYQFLQKVVS
PNPKGEWVSPERQDALANRONNRAVGEDAKNRKPTVTSATSAISKQOQKDLGYKNI
PRNAKAKILRFTYLTQNERQYQLLLELEKELKYLKCKQKLTDLFKDEKDDTD
TTTSTSTSTSSSSSSSSSSSSSSDEGVTSTTSEATATATATTTTTTSTST
STTTTAVNSADEATSVEEHEKDVSESTSGKGTADSQINVAEPISSGVLEPR
TTDQSGSKSGVPTDEQEEKSDVKVPPESEKKEVEAEAGDAEBETSQSAAEESA
ERTSTPETSEPESEDESDIDPSKAPVPEPSRKERTGIFALMKSPSTSSSTOKSKT
AAPSNEVAPENPELVKTKHGYLSKAVDKINYDEKIHQAKLADIRAKEKDYDAK
NKEYEKQLQDQIDIEINSMKAMREETSEKIEVSKNRLVKLIIDVNAHNKKLMI
LKDTENNKQKQEKNEVLDDKQTNVSEIDDLNNEKTNVQKFNDFNTNLNSLSQQLD
AQIFKINQNLKQKQVNEIDNLEKKEDLVQTEENKHLKHNQVLESVENKYLPL
QINDIDNOISLLNEVTIIKQENANEKTLQSLAIKRLDEERRAHEEQLEAEERKKE
EENLLEKQCELEQEAQALDHEQOITQVKYNDQLTQLQKLAETEKELEAVKRE
RFLQAEKALEEQTQKNDAEALQKEILSRQKQAEIGHAENHKIPNDRSQKNTSVL
PKDDSLYEHTEEDVMFA"
complement(3541..3551)
/notes="ARS-consensus"
misc_feature
/notes="ARS-consensus"
ORIGIN

Query Match          91.6%; Score 17.4; DB 4; Length 3678;
Best Local Similarity 94.7%; Pred. No. 2.7e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCCTTTG 19
|||||
Db 3064 CTTTACTTCATAGTCCTTTG 3082

RESULT 34
AY800241           59395 bp DNA linear PRI 06-NOV-2004
LOCUS              Homo sapiens CHK2 checkpoint homolog (S. pombe) (CHK2) gene,
DEFINITION         complete cds.
ACCESSION          AY800241
VERSION            AY800241.1 GI:55140703
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
AUTHORS            1 (bases 1 to 59395)
Livingston,R.J., Rieder,M.J., Chung,M.-W., Ritchie,T.K.,
Olson,A.N., Nguyen,C.P., Nguyen,D.A., Poel,C.L., Chambers,S.W.,

ORIGIN
Query Match          91.6%; Score 17.4; DB 4; Length 1744;
Best Local Similarity 94.7%; Pred. No. 2.6e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCCTTTG 19
|||||
Db 267 CTTTACTTCATAGTCCTTTG 285

RESULT 33
SCYGR130C          3678 bp DNA linear PLN 18-APR-2005
LOCUS              S.cerevisiae chromosome VII reading frame ORF YGR130c.
DEFINITION         S.cerevisiae chromosome VII reading frame ORF YGR130c.
ACCESSION          272915 Y13135
VERSION            272915.1 GI:1323214
KEYWORDS
SOURCE
ORGANISM
Saccharomyces cerevisiae (baker's yeast)
Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE
AUTHORS            van Dyck,L., Skala,J., de Wergifosse,P., Purnelle,B., Talla,E.,
Nawrocki,A., Del Bino,S. and Goffeau,A.
JOURNAL            Unpublished
REFERENCE          2 (bases 1 to 1744)
MIPS
Direct Submission
Submitted (14-MAY-1996) Data collected by MIPS on behalf of the
European yeast chromosome VII sequencing project. MIPS at the
Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152
Martinsried, FRG; E-mail: Mewes@mips.embnnet.org
LOCATION/Qualifiers
1. .1744
/organism="Saccharomyces cerevisiae"
/mol_type="genomic DNA"
/db_xref="taxon:4932"
/chromosome="VII"
complement(744..754)
/notes="ARS-consensus"
misc_feature
882..1406
/notes="unnamed protein product; ORF YGR131w"
codon_start=1
/protein_id="CAA97144.1"
/db_xref="GI:1323217"
/db_xref="GOA:P53279"
/db_xref="SGD:S0003363"
/translation="MLFNTNRQDDPPTOLINOSANTONQQAHQESPYQFLQKVVS
PNPKGEWVSPERQDALANRONNRAVGEDAKNRKPTVTSATSAISKQOQKDLGYKNI
PRNAKAKILRFTYLTQNERQYQLLLELEKELKYLKCKQKLTDLFKDEKDDTD
TTTSTSTSTSSSSSSSSSSSSSSDEGVTSTTSEATATATATTTTTTSTST
STTTTAVNSADEATSVEEHEKDVSESTSGKGTADSQINVAEPISSGVLEPR
TTDQSGSKSGVPTDEQEEKSDVKVPPESEKKEVEAEAGDAEBETSQSAAEESA
ERTSTPETSEPESEDESDIDPSKAPVPEPSRKERTGIFALMKSPSTSSSTOKSKT
AAPSNEVAPENPELVKTKHGYLSKAVDKINYDEKIHQAKLADIRAKEKDYDAK
NKEYEKQLQDQIDIEINSMKAMREETSEKIEVSKNRLVKLIIDVNAHNKKLMI
LKDTENNKQKQEKNEVLDDKQTNVSEIDDLNNEKTNVQKFNDFNTNLNSLSQQLD
AQIFKINQNLKQKQVNEIDNLEKKEDLVQTEENKHLKHNQVLESVENKYLPL
QINDIDNOISLLNEVTIIKQENANEKTLQSLAIKRLDEERRAHEEQLEAEERKKE
EENLLEKQCELEQEAQALDHEQOITQVKYNDQLTQLQKLAETEKELEAVKRE
RFLQAEKALEEQTQKNDAEALQKEILSRQKQAEIGHAENHKIPNDRSQKNTSVL
PKDDSLYEHTEEDVMFA"
complement(3541..3551)
/notes="ARS-consensus"
misc_feature
/notes="ARS-consensus"
ORIGIN

```

Schackwitz,W.S., Sherwood,J.K., Sherwood,A.M., Leithauser,B.J. and Nickerson,D.A.  
TITLE Direct Submission  
JOURNAL Submitted (28-OCT-2004) Genome Sciences, University of Washington, 1705 NE Pacific, Seattle, WA 98195, USA  
COMMENT To cite this work please use: NIEHS-SNPs, Environmental Genome Project, NIEHS ES15478, Department of Genome Sciences, Seattle, WA (URL: <http://egp.gs.washington.edu>).

FEATURES	Location/Qualifiers		
source	1..59195		
	/organism="Homo sapiens"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:9606"		
variation	119		
	/frequency="0.01"		
	/replace="g"		
variation	132		
	/frequency="0.01"		
	/replace="c"		
repeat_region	337..500		
	/rpt_family="MIR"		
	/rpt_type=dispersed		
variation	453		
	/frequency="0.01"		
	/replace="c"		
variation	499		
	/frequency="0.01"		
	/replace="c"		
repeat_region	503..793		
	/rpt_family="Alu"		
	/rpt_type=dispersed		
variation	693..721		
	/frequency="0.46"		
	/replace="a"		
variation	808		
	/frequency="0.11"		
	/replace="a"		
repeat_region	1012..1260		
	/rpt_family="Alu"		
	/rpt_type=dispersed		
variation	1226		
	/frequency="0.47"		
	/replace="t"		
repeat_region	1494..1791		
	/rpt_family="Alu"		
	/rpt_type=dispersed		
variation	1520		
	/frequency="0.07"		
	/replace="t"		
variation	1555		
	/frequency="0.01"		
	/replace="a"		
variation	1767		
	/frequency="0.01"		
	/replace="t"		
variation	2230		
	/frequency="0.01"		
	/replace="g"		
repeat_region	2254..2510		
	/rpt_family="L2"		
	/rpt_type=dispersed		
repeat_region	2583..3015		
	/rpt_family="ERV1"		
	/rpt_type=dispersed		
variation	3065		
	/frequency="0.01"		
	/replace="a"		
variation	3334		
	/frequency="0.01"		
	/replace="t"		
variation	3434..3449		
	/frequency="0.99"		
	/replace="a"		
variation	3487		
	/frequency="0.01"		
	/replace="a"		
variation	3764		
	/frequency="0.01"		
	/replace="c"		
variation	3772		
	/frequency="0.01"		
	/replace="g"		
variation	4121		
	/frequency="0.39"		
	/replace="c"		
variation	4195		
	/frequency="0.12"		
	/replace="a"		
variation	4227		
	/frequency="0.01"		
	/replace="a"		
gene	4243..58366		
	/gene="CHEK2"		
mRNA	join(4243..4308,11351..11675,20712..20836,20955..21102,26594..28684,34085..34193,36045..36098,42538..42599,46167..46266,49118..49204,50232..50395,50863..50978,51988..52073,56890..56970,58119..58366)		
	/gene="CHEK2"		
	/product="CHK2 checkpoint homolog (S. pombe)"		
variation	4555		
	/gene="CHEK2"		
	/frequency="0.01"		
	/replace="g"		
variation	4594		
	/gene="CHEK2"		
	/frequency="0.01"		
	/replace="t"		
variation	4708		
	/gene="CHEK2"		
	/frequency="0.01"		
	/replace="t"		
variation	4746		
	/gene="CHEK2"		
	/frequency="0.01"		
	/replace="g"		
variation	4788		
	/gene="CHEK2"		
	/frequency="0.01"		
	/replace="a"		
variation	4902		
	/gene="CHEK2"		
	/frequency="0.02"		
	/replace="g"		
repeat_region	4905..5203		
	/rpt_family="Alu"		
	/rpt_type=dispersed		
variation	4908		
	/gene="CHEK2"		
	/frequency="0.48"		
	/replace="c"		
variation	4983		
	/gene="CHEK2"		
	/frequency="0.01"		
	/replace="g"		
repeat_region	5204..5272		
	/rpt_family="L1"		
	/rpt_type=dispersed		
repeat_region	5276..5869		
	/rpt_family="ERVK"		
	/rpt_type=dispersed		
variation	5310		
	/gene="CHEK2"		
	/frequency="0.01"		
	/replace="a"		
variation	5413		
	/gene="CHEK2"		

```

/replace="0.01"
5414
variation
/replace="t"
/replace="a"
5810
variation
/replace="a"
/replace="g"
5832
variation
/replace="a"
/replace="t"
5870.6157
repeat_region
/replace="t"
/replace="g"
5878
variation
/replace="t"
/replace="g"
5915
variation
/replace="t"
/replace="t"
5984
variation
/replace="t"
/replace="t"
6158.6274
repeat_region
/replace="t"
/replace="t"
6177
variation
/replace="t"
/replace="g"
6224.8984
misc_feature
/replace="t"
/replace="g"
6275.6365
repeat_region
/replace="t"
/replace="g"
6275.6365

```

```

Query Match 91.6%; Score 17.4; DB 5; Length 59395;
Best Local Similarity 94.7%; Pred. No. 3.4e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19
Db 30742 CTTTACTTCATAGTCTTTG 30760

RESULT 35
AC157840/c 108957 bp DNA linear PRI 26-JAN-2006
LOCUS AC157840
DEFINITION Aotus nancymaae clone CH258-147H21, complete sequence.
ACCESSION AC157840
VERSION AC157840.3 GI:85719332
KEYWORDS HTG.
SOURCE Aotus nancymaae (Ma's night monkey)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Platyrrhini;
Cebidae; Aotinae; Aotus.
1 (bases 1 to 108957)
Antoneilles,A., Bass,D., Benjamin,B., Bera,J., Blakesley,R.W.,
Bouffard,G.G., Brinkley,C., Brooks,S., Chu,G., Coleman,H.,
Franks,S., Fukenko,T., Gestole,M., Greene,A., Guan,X., Gupta,J.,
Hunter,G., Hurle,B., Idol,J.R., Kwong,P., Latic,P., Larson,S.,
Lee-Tin,S.-Q., Legaspi,R., Madden,M., Maduro,Q.L., Maduro,V.B.,
Margulies,E.H., Masello,C., Maskeri,B., McDowell,J.,
Montemayor,C., Mullikin,J.C., Park,M., Prasad,A., Puri,O.,
Rantz,K., Reddix-Dugue,N., Sante,A., Schandler,K., Schueler,M.G.,
Sison,C., Stantripop,S., Taye,A., Thomas,J.W., Thomas,P.J.,

```

```

Teipouri,V., Ung,L., Vogt,J.L., Wetherby,K.D., Withers,T.R.,
Young,A. and Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
REFERENCE 2 (bases 1 to 108957)
AUTHORS Green,E.D.
TITLE Direct Submission
JOURNAL Submitted (02-MAR-2005) NIH Intramural Sequencing Center, 5625
Fishers Lane, Rockville, MD 20852, USA
REFERENCE 3 (bases 1 to 108957)
AUTHORS Green,E.D.
TITLE Direct Submission
JOURNAL Submitted (07-APR-2005) NIH Intramural Sequencing Center, 5625
Fishers Lane, Rockville, MD 20852, USA
REFERENCE 4 (bases 1 to 108957)
AUTHORS Green,E.D.
TITLE Direct Submission
JOURNAL Submitted (26-JAN-2006) NIH Intramural Sequencing Center, 5625
Fishers Lane, Rockville, MD 20852, USA
COMMENT On Jan 26, 2006 this sequence version replaced gi:62358351.
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_zoo@nhgri.nih.gov
----- Project Information
Center project name: hzv
Center clone name: 147H21

```

This sequence was finished as follows unless otherwise noted:  
all regions were double-stranded, sequenced with an  
alternate chemistry, or covered by high quality data  
(i.e., phred quality >= 30); an attempt was made to resolve  
all sequencing problems, such as compressions and repeats;  
all regions were covered by at least one plasmid subclone  
or more than one M13 subclone; and the assembly was confirmed  
by restriction digest.

CLONE LENGTH: This sequence represents the entire insert of  
this clone unless otherwise noted. If there are overlapping  
clones, the overlaps are noted in the beginning and end of  
the Features section.

FEATURES  
source

```

1.108957 Location/Qualifiers
/organism="Aotus nancymaae"
/mol_type="genomic DNA"
/db_xref="taxon:37293"
/clone="CH258-147H21"
/clone_lib="CH258"
/notes="BAC resource: http://bacpac.chori.org/"
1.97928

```

misc\_feature

```

/notes="clone overlaps with GenBank Accession Number
AC151368.3 (nucleotides 85363-189563) clone CH258-251019
(center project name hzf)"
97935..>108957

```

misc\_feature

```

/notes="this sequence is not the entire insert of clone
CH258-147H21; clone overlaps with GenBank Accession Number
AC151487.3 (nucleotides 1-56127) clone CH258-321B24
(center project name h21); this annotated segment
represents overlap with nucleotides 1-11023 of AC151487.3"

```

ORIGIN

```

Query Match 91.6%; Score 17.4; DB 5; Length 108957;
Best Local Similarity 94.7%; Pred. No. 3.4e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19
Db 96005 CTTTACTTCATAGTCTTG 95987

RESULT 36
AP008215_067/c

```

## WPCOMMENT

Sequence split into 227 fragments LOCUS AP008215 Accession AP008215

Fragment Name	Begin	End
AP008215_000	1	110000
AP008215_001	100001	210000
AP008215_002	200001	310000
AP008215_003	300001	410000
AP008215_004	400001	510000
AP008215_005	500001	610000
AP008215_006	600001	710000
AP008215_007	700001	810000
AP008215_008	800001	910000
AP008215_009	900001	1010000
AP008215_010	1100001	1110000
AP008215_011	1100001	1210000
AP008215_012	1200001	1310000
AP008215_013	1300001	1410000
AP008215_014	1400001	1510000
AP008215_015	1500001	1610000
AP008215_016	1600001	1710000
AP008215_017	1700001	1810000
AP008215_018	1800001	1910000
AP008215_019	1900001	2010000
AP008215_020	2000001	2110000
AP008215_021	2100001	2210000
AP008215_022	2200001	2310000
AP008215_023	2300001	2410000
AP008215_024	2400001	2510000
AP008215_025	2500001	2610000
AP008215_026	2600001	2710000
AP008215_027	2700001	2810000
AP008215_028	2800001	2910000
AP008215_029	2900001	3010000
AP008215_030	3000001	3110000
AP008215_031	3100001	3210000
AP008215_032	3200001	3310000
AP008215_033	3300001	3410000
AP008215_034	3400001	3510000
AP008215_035	3500001	3610000
AP008215_036	3600001	3710000
AP008215_037	3700001	3810000
AP008215_038	3800001	3910000
AP008215_039	3900001	4010000
AP008215_040	4000001	4110000
AP008215_041	4100001	4210000
AP008215_042	4200001	4310000
AP008215_043	4300001	4410000
AP008215_044	4400001	4510000
AP008215_045	4500001	4610000
AP008215_046	4600001	4710000
AP008215_047	4700001	4810000
AP008215_048	4800001	4910000
AP008215_049	4900001	5010000
AP008215_050	5000001	5110000
AP008215_051	5100001	5210000
AP008215_052	5200001	5310000
AP008215_053	5300001	5410000
AP008215_054	5400001	5510000
AP008215_055	5500001	5610000
AP008215_056	5600001	5710000
AP008215_057	5700001	5810000
AP008215_058	5800001	5910000
AP008215_059	5900001	6010000
AP008215_060	6000001	6110000
AP008215_061	6100001	6210000
AP008215_062	6200001	6310000
AP008215_063	6300001	6410000
AP008215_064	6400001	6510000
AP008215_065	6500001	6610000
AP008215_066	6600001	6710000
AP008215_067	6700001	6810000
AP008215_068	6800001	6910000
AP008215_069	6900001	7010000
AP008215_070	7000001	7110000
AP008215_071	7100001	7210000
AP008215_072	7200001	7310000
AP008215_073	7300001	7410000
AP008215_074	7400001	7510000
AP008215_075	7500001	7610000
AP008215_076	7600001	7710000
AP008215_077	7700001	7810000
AP008215_078	7800001	7910000
AP008215_079	7900001	8010000
AP008215_080	8000001	8110000
AP008215_081	8100001	8210000
AP008215_082	8200001	8310000
AP008215_083	8300001	8410000
AP008215_084	8400001	8510000
AP008215_085	8500001	8610000
AP008215_086	8600001	8710000
AP008215_087	8700001	8810000
AP008215_088	8800001	8910000
AP008215_089	8900001	9010000
AP008215_090	9000001	9110000
AP008215_091	9100001	9210000
AP008215_092	9200001	9310000
AP008215_093	9300001	9410000
AP008215_094	9400001	9510000
AP008215_095	9500001	9610000
AP008215_096	9600001	9710000
AP008215_097	9700001	9810000
AP008215_098	9800001	9910000
AP008215_099	9900001	10010000
AP008215_100	10000001	10110000
AP008215_101	10100001	10210000
AP008215_102	10200001	10310000
AP008215_103	10300001	10410000
AP008215_104	10400001	10510000
AP008215_105	10500001	10610000
AP008215_106	10600001	10710000
AP008215_107	10700001	10810000
AP008215_108	10800001	10910000
AP008215_109	10900001	11010000
AP008215_110	11000001	11110000
AP008215_111	11100001	11210000
AP008215_112	11200001	11310000
AP008215_113	11300001	11410000
AP008215_114	11400001	11510000
AP008215_115	11500001	11610000
AP008215_116	11600001	11710000
AP008215_117	11700001	11810000
AP008215_118	11800001	11910000
AP008215_119	11900001	12010000
AP008215_120	12000001	12110000
AP008215_121	12100001	12210000
AP008215_122	12200001	12310000
AP008215_123	12300001	12410000
AP008215_124	12400001	12510000
AP008215_125	12500001	12610000
AP008215_126	12600001	12710000
AP008215_127	12700001	12810000
AP008215_128	12800001	12910000
AP008215_129	12900001	13010000
AP008215_130	13000001	13110000
AP008215_131	13100001	13210000
AP008215_132	13200001	13310000
AP008215_133	13300001	13410000
AP008215_134	13400001	13510000
AP008215_135	13500001	13610000
AP008215_136	13600001	13710000
AP008215_137	13700001	13810000
AP008215_138	13800001	13910000
AP008215_139	13900001	14010000
AP008215_140	14000001	14110000
AP008215_141	14100001	14210000
AP008215_142	14200001	14310000

```
AP008215_143 14300001 14410000
AP008215_144 14400001 14510000
AP008215_145 14500001 14610000
AP008215_146 14600001 14710000
AP008215_147 14700001 14810000
AP008215_148 14800001 14910000
AP008215_149 14900001 15010000
AP008215_150 15000001 15110000
AP008215_151 15100001 15210000
AP008215_152 15200001 15310000
AP008215_153 15300001 15410000
AP008215_154 15400001 15510000
AP008215_155 15500001 15610000
AP008215_156 15600001 15710000
AP008215_157 15700001 15810000
AP008215_158 15800001 15910000
AP008215_159 15900001 16010000
AP008215_160 16000001 16110000
AP008215_161 16100001 16210000
AP008215_162 16200001 16310000
AP008215_163 16300001 16410000
AP008215_164 16400001 16510000
AP008215_165 16500001 16610000
AP008215_166 16600001 16710000
AP008215_167 16700001 16810000
AP008215_168 16800001 16910000
AP008215_169 16900001 17010000
AP008215_170 17000001 17110000
AP008215_171 17100001 17210000
AP008215_172 17200001 17310000
AP008215_173 17300001 17410000
AP008215_174 17400001 17510000
AP008215_175 17500001 17610000
AP008215_176 17600001 17710000
AP008215_177 17700001 17810000
AP008215_178 17800001 17910000
AP008215_179 17900001 18010000
AP008215_180 18000001 18110000
AP008215_181 18100001 18210000
AP008215_182 18200001 18310000
AP008215_183 18300001 18410000
AP008215_184 18400001 18510000
AP008215_185 18500001 18610000
AP008215_186 18600001 18710000
AP008215_187 18700001 18810000
AP008215_188 18800001 18910000
AP008215_189 18900001 19010000
AP008215_190 19000001 19110000
AP008215_191 19100001 19210000
AP008215_192 19200001 19310000
AP008215_193 19300001 19410000
AP008215_194 19400001 19510000
AP008215_195 19500001 19610000
AP008215_196 19600001 19710000

Query Match          91.6%; Score 17.4; DB 4; Length 110000;
Best Local Similarity 94.7%; Pred. No. 3.4e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGCTTTG 19
    |||||
Db 67251 CTTTACTTCATAGCTTCG 67233

RESULT 37
AP008207_149
WPCOMMENT
Sequence split into 433 fragments LOCUS AP008207 Accession AP008207
Fragment Name      Begin      End
AP008207_000      1      110000
AP008207_001     100001     210000
AP008207_002     200001     310000
AP008207_003     300001     410000
AP008207_004     400001     510000
AP008207_005     500001     610000
AP008207_006     600001     710000
AP008207_007     700001     810000
AP008207_008     800001     910000
AP008207_009     900001    1010000
AP008207_010    1000001    1110000
AP008207_011    1100001    1210000
AP008207_012    1200001    1310000
AP008207_013    1300001    1410000
AP008207_014    1400001    1510000
AP008207_015    1500001    1610000
AP008207_016    1600001    1710000
AP008207_017    1700001    1810000
AP008207_018    1800001    1910000
AP008207_019    1900001    2010000
AP008207_020    2000001    2110000
AP008207_021    2100001    2210000
AP008207_022    2200001    2310000
AP008207_023    2300001    2410000
AP008207_024    2400001    2510000
AP008207_025    2500001    2610000
AP008207_026    2600001    2710000
AP008207_027    2700001    2810000
AP008207_028    2800001    2910000
AP008207_029    2900001    3010000
AP008207_030    3000001    3110000
AP008207_031    3100001    3210000
AP008207_032    3200001    3310000
AP008207_033    3300001    3410000
AP008207_034    3400001    3510000
AP008207_035    3500001    3610000
AP008207_036    3600001    3710000
AP008207_037    3700001    3810000
AP008207_038    3800001    3910000
AP008207_039    3900001    4010000
AP008207_040    4000001    4110000
AP008207_041    4100001    4210000
AP008207_042    4200001    4310000
AP008207_043    4300001    4410000
AP008207_044    4400001    4510000
AP008207_045    4500001    4610000
AP008207_046    4600001    4710000
AP008207_047    4700001    4810000
AP008207_048    4800001    4910000
AP008207_049    4900001    5010000
AP008207_050    5000001    5110000
AP008207_051    5100001    5210000
AP008207_052    5200001    5310000
AP008207_053    5300001    5410000
AP008207_054    5400001    5510000
AP008207_055    5500001    5610000
AP008207_056    5600001    5710000
AP008207_057    5700001    5810000
AP008207_058    5800001    5910000
AP008207_059    5900001    6010000
AP008207_060    6000001    6110000
AP008207_061    6100001    6210000
AP008207_062    6200001    6310000
AP008207_063    6300001    6410000
AP008207_064    6400001    6510000
AP008207_065    6500001    6610000
AP008207_066    6600001    6710000
AP008207_067    6700001    6810000
AP008207_068    6800001    6910000
AP008207_069    6900001    7010000
AP008207_070    7000001    7110000
AP008207_071    7100001    7210000
AP008207_072    7200001    7310000
AP008207_073    7300001    7410000
AP008207_074    7400001    7510000
AP008207_075    7500001    7610000
AP008207_076    7600001    7710000
```



AP008207\_077 7700001 7810000  
AP008207\_078 7800001 7910000  
AP008207\_079 7900001 8010000  
AP008207\_080 8000001 8110000  
AP008207\_081 8100001 8210000  
AP008207\_082 8200001 8310000  
AP008207\_083 8300001 8410000  
AP008207\_084 8400001 8510000  
AP008207\_085 8500001 8610000  
AP008207\_086 8600001 8710000  
AP008207\_087 8700001 8810000  
AP008207\_088 8800001 8910000  
AP008207\_089 8900001 9010000  
AP008207\_090 9000001 9110000  
AP008207\_091 9100001 9210000  
AP008207\_092 9200001 9310000  
AP008207\_093 9300001 9410000  
AP008207\_094 9400001 9510000  
AP008207\_095 9500001 9610000  
AP008207\_096 9600001 9710000  
AP008207\_097 9700001 9810000  
AP008207\_098 9800001 9910000  
AP008207\_099 9900001 10010000  
AP008207\_100 1000001 10110000  
AP008207\_101 10100001 10210000  
AP008207\_102 10200001 10310000  
AP008207\_103 10300001 10410000  
AP008207\_104 10400001 10510000  
AP008207\_105 10500001 10610000  
AP008207\_106 10600001 10710000  
AP008207\_107 10700001 10810000  
AP008207\_108 10800001 10910000  
AP008207\_109 10900001 11010000  
AP008207\_110 11000001 11110000  
AP008207\_111 11100001 11210000  
AP008207\_112 11200001 11310000  
AP008207\_113 11300001 11410000  
AP008207\_114 11400001 11510000  
AP008207\_115 11500001 11610000  
AP008207\_116 11600001 11710000  
AP008207\_117 11700001 11810000  
AP008207\_118 11800001 11910000  
AP008207\_119 11900001 12010000  
AP008207\_120 12000001 12110000  
AP008207\_121 12100001 12210000  
AP008207\_122 12200001 12310000  
AP008207\_123 12300001 12410000  
AP008207\_124 12400001 12510000  
AP008207\_125 12500001 12610000  
AP008207\_126 12600001 12710000  
AP008207\_127 12700001 12810000  
AP008207\_128 12800001 12910000  
AP008207\_129 12900001 13010000  
AP008207\_130 13000001 13110000  
AP008207\_131 13100001 13210000  
AP008207\_132 13200001 13310000  
AP008207\_133 13300001 13410000  
AP008207\_134 13400001 13510000  
AP008207\_135 13500001 13610000  
AP008207\_136 13600001 13710000  
AP008207\_137 13700001 13810000  
AP008207\_138 13800001 13910000  
AP008207\_139 13900001 14010000  
AP008207\_140 14000001 14110000  
AP008207\_141 14100001 14210000  
AP008207\_142 14200001 14310000  
AP008207\_143 14300001 14410000  
AP008207\_144 14400001 14510000  
AP008207\_145 14500001 14610000  
AP008207\_146 14600001 14710000  
AP008207\_147 14700001 14810000  
AP008207\_148 14800001 14910000  
AP008207\_149 14900001 15010000

AP008207\_150 15000001 15110000  
AP008207\_151 15100001 15210000  
AP008207\_152 15200001 15310000  
AP008207\_153 15300001 15410000  
AP008207\_154 15400001 15510000  
AP008207\_155 15500001 15610000  
AP008207\_156 15600001 15710000  
AP008207\_157 15700001 15810000  
AP008207\_158 15800001 15910000  
AP008207\_159 15900001 16010000  
AP008207\_160 16000001 16110000  
AP008207\_161 16100001 16210000  
AP008207\_162 16200001 16310000  
AP008207\_163 16300001 16410000  
AP008207\_164 16400001 16510000  
AP008207\_165 16500001 16610000  
AP008207\_166 16600001 16710000  
AP008207\_167 16700001 16810000  
AP008207\_168 16800001 16910000  
AP008207\_169 16900001 17010000  
AP008207\_170 17000001 17110000  
AP008207\_171 17100001 17210000  
AP008207\_172 17200001 17310000  
AP008207\_173 17300001 17410000  
AP008207\_174 17400001 17510000  
AP008207\_175 17500001 17610000  
AP008207\_176 17600001 17710000  
AP008207\_177 17700001 17810000  
AP008207\_178 17800001 17910000  
AP008207\_179 17900001 18010000  
AP008207\_180 18000001 18110000  
AP008207\_181 18100001 18210000  
AP008207\_182 18200001 18310000  
AP008207\_183 18300001 18410000  
AP008207\_184 18400001 18510000  
AP008207\_185 18500001 18610000  
AP008207\_186 18600001 18710000  
AP008207\_187 18700001 18810000  
AP008207\_188 18800001 18910000  
AP008207\_189 18900001 19010000  
AP008207\_190 19000001 19110000  
AP008207\_191 19100001 19210000  
AP008207\_192 19200001 19310000  
AP008207\_193 19300001 19410000  
AP008207\_194 19400001 19510000  
AP008207\_195 19500001 19610000  
AP008207\_196 19600001 19710000

Query Match 91.6%; Score 17.4; DB 4; Length 110000;  
Best Local Similarity 94.7%; Pred. No. 3.4e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Cy 1 CTTTACTTCATAGTCTTTG 19  
||| ||||| ||||| |||||  
Db 20335 CTTTGCTTCATAGTCTTTG 20353

RESULT 38  
CP000155\_24/c  
WPCOMMENT

Sequence split into 73 fragments LOCUS CP000155 Accession CP000155

Fragment Name	Begin	End
CP000155_00	1	110000
CP000155_01	100001	210000
CP000155_02	200001	310000
CP000155_03	300001	410000
CP000155_04	400001	510000
CP000155_05	500001	610000
CP000155_06	600001	710000
CP000155_07	700001	810000
CP000155_08	800001	910000
CP000155_09	900001	1010000
CP000155_10	1000001	1110000

```
CP000155-11 1100001 1210000
CP000155-12 1200001 1310000
CP000155-13 1300001 1410000
CP000155-14 1400001 1510000
CP000155-15 1500001 1610000
CP000155-16 1600001 1710000
CP000155-17 1700001 1810000
CP000155-18 1800001 1910000
CP000155-19 1900001 2010000
CP000155-20 2000001 2110000
CP000155-21 2100001 2210000
CP000155-22 2200001 2310000
CP000155-23 2300001 2410000
CP000155-24 2400001 2510000
CP000155-25 2500001 2610000
CP000155-26 2600001 2710000
CP000155-27 2700001 2810000
CP000155-28 2800001 2910000
CP000155-29 2900001 3010000
CP000155-30 3000001 3110000
CP000155-31 3100001 3210000
CP000155-32 3200001 3310000
CP000155-33 3300001 3410000
CP000155-34 3400001 3510000
CP000155-35 3500001 3610000
CP000155-36 3600001 3710000
CP000155-37 3700001 3810000
CP000155-38 3800001 3910000
CP000155-39 3900001 4010000
CP000155-40 4000001 4110000
CP000155-41 4100001 4210000
CP000155-42 4200001 4310000
CP000155-43 4300001 4410000
CP000155-44 4400001 4510000
CP000155-45 4500001 4610000
CP000155-46 4600001 4710000
CP000155-47 4700001 4810000
CP000155-48 4800001 4910000
CP000155-49 4900001 5010000
CP000155-50 5000001 5110000
CP000155-51 5100001 5210000
CP000155-52 5200001 5310000
CP000155-53 5300001 5410000
CP000155-54 5400001 5510000
CP000155-55 5500001 5610000
CP000155-56 5600001 5710000
CP000155-57 5700001 5810000
CP000155-58 5800001 5910000
CP000155-59 5900001 6010000
CP000155-60 6000001 6110000
CP000155-61 6100001 6210000
CP000155-62 6200001 6310000
CP000155-63 6300001 6410000
CP000155-64 6400001 6510000
CP000155-65 6500001 6610000
CP000155-66 6600001 6710000
CP000155-67 6700001 6810000
CP000155-68 6800001 6910000
CP000155-69 6900001 7010000
CP000155-70 7000001 7110000
CP000155-71 7100001 7210000
CP000155-72 7200001 7215267

Continuation (25 of 73) of CP000155 from base 2400001 (CP000155 Hahella chejuensis KCTC

Query Match 91.6%; Score 17.4; DB 15; Length 110000;
Best Local Similarity 94.7%; Pred. No. 3.4e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGCTTTG 19
| | | | | | | | | | | | | | | | | | | | |
Db 60999 CTTTACTTCATAGCTTTG 60881
```

```
RESULT 39
AP001818/c
LOCUS
DEFINITION Homo sapiens genomic DNA, chromosome 4q22-q24, clone:2222M9,
complete sequence.
ACCESSION AP001818
VERSION AP001818.2 GI:10998160
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE
AUTHORS Tsai,S.F.
TITLE Genome Sequencing of the Chromosome 4q Region Implicated in Human
Hepatocellular Carcinoma Pathogenesis 1
JOURNAL Unpublished
REFERENCE
AUTHORS Tsai,S.F.
TITLE Direct Submission
JOURNAL Submitted (14-APR-2000) Shih-Feng Tsai, National Yang-Ming
University, Institute of Genetics; 155 Li-Rong St. Section 2,
Peitou, Taipei, Taiwan 11221, Republic of China
(Pe-mail:ympetsai@ym.edu.tw, URL:http://genome.ym.edu.tw/,
Tel:886-2-28267043, Fax:886-2-28264930)
COMMENT
Project AP
On Oct 24, 2000 this sequence version replaced gi:7594883.
Quality: the expected Phred/Phrap calculated error rate (per 10 kb)
is 0.0007; Estimated total number of errors is 0.0087.
FEATURES
source
1. .116321
/organism="Homo sapiens"
/mol type="genomic DNA"
/db xref="taxon:9606"
/chromosome="4"
/map="4q22-q24"
/clone="2222M9"
ORIGIN
Query Match 91.6%; Score 17.4; DB 5; Length 116321;
Best Local Similarity 94.7%; Pred. No. 3.4e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGCTTTG 19
| | | | | | | | | | | | | | | | | | | | |
Db 79498 CATTACTTCATAGCTTTG 79480

RESULT 40
AC150029/c
LOCUS
DEFINITION Sorex araneus clone SA_Ba-362A8, WORKING DRAFT SEQUENCE, 2 ordered
pieces.
ACCESSION AC150029
VERSION AC150029.2 GI:50979356
KEYWORDS HTG; HTGS PHASE2; HTGS DRAFT.
SOURCE Sorex araneus (European shrew)
ORGANISM Sorex araneus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Insectivora; Soricidae;
Soricinae; Sorex.
REFERENCE
AUTHORS 1 (bases 1 to 121462)
Antonellis,A., Ayele,K., Benjamin,B., Blakesley,R.W.,
Bouffard,G.G., Brinkley,C., Brooks,S., Chu,G., Coleman,B.,
Coleman,H., Daki,N., Engle,J., Guan,X., Gupta,J., Haghighi,P.,
Han,J., Hansen,N., Ho,S.-L., Hu,P., Hurlb,B., Idol,J.R., Jones,C.,
Karlsn,B., Kim,H., Kwong,P., Laric,P., Larson,S., Lee-lin,S.-Q.,
Legaspi,R., Madden,M., Maduro,Q.I., Maduro,V.B., Margulies,E.H.,
Masello,C., Maskeri,B., McDowell,J., Mullikin,J.C., Pauguirgan,C.,
Park,M., Portnoy,M.E., Prasad,A., Puri,O., Reddix-Dugue,N.,
Schandler,K., Schueler,M.G., Shah,K., Sison,C., Stantriop,S.,
Thomas,J.W., Thomas,P.J., Tsipouri,V., Vogt,J.L., Wetherby,K.D.,
```

```

Young, A. and Green, E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 121462)
Direct Submission
Submitted (30-JUN-2004) NIH Intramural Sequencing Center, 8717
Grovermont Circle, Gaithersburg, MD 20877, USA
3 (bases 1 to 121462)
Green, E.D.
Direct Submission
Submitted (05-AUG-2004) NIH Intramural Sequencing Center, 8717
Grovermont Circle, Gaithersburg, MD 20877, USA
On Aug 5, 2004 this sequence version replaced gi:49457917.
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc.zoo@nhgri.nih.gov
----- Project Information
Center project name: ggc
Center clone name: 362A08

The sequence data in this record represents an 'enhanced'
version of a Phase 2 submission. Specifically, the indicated
order and orientation of each sequence contig has been
established using one or more of the following: read-pair
data from individual subclones, overlaps with neighboring
clones, alignment with available reference sequence (e.g.,
human), and/or confirmation by PCR testing. In addition,
the sequence assembly is based on at least 8X average
coverage in Q20 bases and has been reviewed to rule out
gross misassemblies, the low-quality ends of sequence
contigs have been trimmed away, and each base is associated
with a Phrap-derived quality score.
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 121178 bases at least Q40
Consensus quality: 121304 bases at least Q30
Consensus quality: 121341 bases at least Q20
Insert size: 125000; agarose-fp
Insert size: 121362; sum-of-contigs
Quality coverage: 8.26x in Q20 bases; agarose-fp
Quality coverage: 8.51x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved
* 1 25692: contig of 25692 bp in length
* 25693 25792: gap of unknown length
* 25793 121462: contig of 95670 bp in length.
Location/Qualifiers
1. 121462
/organism="Sorex araneus"
/mol_type="genomic DNA"
/db_xref="taxon:42254"
/clone="SA Ba-362A8"
/clone_lib="SA_Ba"
misc_feature 1. 25692
/note="assembly_fragment
clone_end:r7
vector_side:left"
25693. 25792
/estimated_length=unknown
misc_feature 25793. 121462

```

/Note="assembly\_fragment  
clone\_end:SP6  
vector\_side:right"

Query Match 91.6%; Score 17.4; DB 12; Length 121462;  
Best Local Similarity 94.7%; Pred. No. 3.4e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19  
|||||  
Db 51546 CTTTACTTCATAGTATTG 51528  
|||||

RESULT 41  
AC167325 121581 bp DNA linear HTG 24-NOV-2005  
LOCUS Loxodonta africana clone VMRC15-443K8, WORKING DRAFT SEQUENCE, 8  
DEFINITION ordered pieces.  
AC167325  
AC167325.2 GI:82654480  
HTG; HTGS\_PHAZE2; HTGS\_DRAFT.  
KEYWORDS Loxodonta africana (African savanna elephant)  
SOURCE Loxodonta africana  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Afrotheria; Proboscidea; Elephantidae;  
Loxodonta.  
REFERENCE 1 (bases 1 to 121581)  
AUTHORS  
Antonellis, A., Ayele, K., Bass, D., Benjamin, B., Bera, J.,  
Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S., Chu, G.,  
Coleman, H., Franks, S., Fuxsenko, T., Gestole, M., Greene, A., Guan, X.,  
Gupta, J., Gurson, N., Haghighi, P., Han, E., Han, J., Hansen, N.,  
Ho, S.-L., Hu, P., Hunter, G., Hurie, B., Idol, J.R., Kwong, P.,  
Laric, P., Larson, S., Lee-Lin, S.-Q., Legaspi, R., Madden, M.,  
Maduro, Q.L., Maduro, V.B., Margulies, E.H., Masello, C., Maskeri, B.,  
McDowell, J., Montemayor, C., Mullikin, J.C., Park, M., Portnoy, M.E.,  
Prasad, A., Puri, O., Rantz, K., Reddix-Dugue, N., Sante, A.,  
Schandler, K., Schueler, M.G., Sison, C., Stantropop, S., Tave, A.,  
Wetherby, K.D., Withers, T.R., Young, A., Ung, L., Vogt, J.L.,  
Thomas, J.W., Thomas, P.J., Tsipouri, V., Ung, L., Vogt, J.L.,  
Green, E.D.  
NISC Comparative Sequencing Initiative  
Unpublished  
REFERENCE 2 (bases 1 to 121581)  
AUTHORS  
Green, E.D.  
Direct Submission  
Submitted (24-AUG-2005) NIH Intramural Sequencing Center, 5625  
Fishers Lane, Rockville, MD 20852, USA  
3 (bases 1 to 121581)  
Green, E.D.  
Direct Submission  
Submitted (24-NOV-2005) NIH Intramural Sequencing Center, 5625  
Fishers Lane, Rockville, MD 20852, USA  
On Nov 24, 2005 this sequence version replaced gi:73745553.  
----- Genome Center  
Center: NIH Intramural Sequencing Center  
Center code: NISC  
Web site: http://www.nisc.nih.gov  
Contact: nisc.zoo@nhgri.nih.gov  
----- Project Information  
Center project name: jrf  
Center clone name: 443K08

The sequence data in this record represents an 'enhanced'  
version of a Phase 2 submission. Specifically, the indicated  
order and orientation of each sequence contig has been  
established using one or more of the following: read-pair  
data from individual subclones, overlaps with neighboring  
clones, alignment with available reference sequence (e.g.,  
human), and/or confirmation by PCR testing. In addition,  
the sequence assembly is based on at least 8X average  
coverage in Q20 bases and has been reviewed to rule out  
gross misassemblies, the low-quality ends of sequence  
contigs have been trimmed away, and each base is associated

FEATURES  
source  
1. 121462  
/organism="Sorex araneus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:42254"  
/clone="SA Ba-362A8"  
/clone\_lib="SA\_Ba"  
misc\_feature 1. 25692  
/note="assembly\_fragment  
clone\_end:r7  
vector\_side:left"  
25693. 25792  
/estimated\_length=unknown  
misc\_feature 25793. 121462

with a Phrap-derived quality score.  
----- Summary Statistics  
Sequencing vector: plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 119217 bases at least Q40  
Consensus quality: 120157 bases at least Q30  
Consensus quality: 120697 bases at least Q20  
Insert size: 148000; agarose-fp  
Quality coverage: 9.21x in Q20 bases; agarose-fp  
Quality coverage: 11.27x in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently consists of 8 contigs. Gaps between the contigs are represented as runs of N's. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

\* This sequence will be replaced

\* By the finished sequence as soon as it is available and the accession number will be preserved.

\* 1 47612: contig of 47612 bp in length

\* 47613 47712: gap of unknown length

\* 47713 56509: contig of 8797 bp in length

\* 56510 56609: gap of unknown length

\* 56610 59676: contig of 3067 bp in length

\* 59677 59776: gap of unknown length

\* 59777 73256: contig of 13480 bp in length

\* 73257 73356: gap of unknown length

\* 73357 82649: contig of 9293 bp in length

\* 82650 82749: gap of unknown length

\* 82750 85544: contig of 2795 bp in length

\* 85544 85644: gap of unknown length

\* 85645 106869: contig of 21225 bp in length

\* 106870 106969: gap of unknown length

\* 106970 121581: contig of 14612 bp in length.

FEATURES  
source  
  
Location/Qualifiers  
organism="Loxodonta africana"  
mol\_type="genomic DNA"  
db\_xref="taxon:9785"  
clone="VMRC15-443K8"  
clone\_lib="VMRC15"  
notes="BAC resource: http://bacpac.chori.org/"  
misc\_feature  
1..47612  
note="assembly\_fragment missing Sp6 clone\_end on 5' of insert"

misc\_feature  
44142..121581  
note="clone overlaps with GenBank Accession Number AC163963 clone VMRC15-243J21 (center project name jrg)"

gap  
47613..47712  
estimated\_length=unknown

misc\_feature  
47713..56509  
note="assembly\_fragment"

gap  
56510..56609  
estimated\_length=unknown

misc\_feature  
56610..59676  
note="assembly\_fragment"

gap  
59677..59776  
estimated\_length=unknown

misc\_feature  
59777..73256  
note="assembly\_fragment"

gap  
73257..73356  
estimated\_length=unknown

misc\_feature  
73357..82649  
note="assembly\_fragment"

gap  
82650..82749  
estimated\_length=unknown

misc\_feature  
82750..85544  
note="assembly\_fragment"

gap  
85545..85644  
estimated\_length=unknown

```
misc_feature
100
/note="Clone_right_end: RP11-329J7"
6496..6642
/note="Single clone region. pUC subcloned to give 5X
coverage over region."
join(complement(37948..88208),
complement(AL023281.1:85702..96147),
complement(AL035397.4:30093..30600),
complement(AL050313.6:101..1143),
complement(AL035453.4:18624..31904),
complement(AL033538.1:17058..69410))
/locus_tag="RP3-477H23.1-001"
join(complement(88018..88208),complement(37948..38226),
complement(AL023281.1:96000..96147),
complement(AL023281.1:87084..87356),
complement(AL023281.1:85702..85832),
complement(AL035397.4:30093..30600),
complement(AL050313.6:101..1143),
complement(AL035453.4:18624..31904),
complement(AL035453.4:29823..30346),
complement(AL035453.4:25715..25824),
complement(AL035453.4:23459..23588),
complement(AL035453.4:20734..20952),
complement(AL035453.4:18624..18789),
complement(AL03538.1:69270..69410),
complement(AL03538.1:53292..53436),
complement(AL03538.1:40418..40597),
complement(AL03538.1:37584..38304),
complement(AL03538.1:35179..35303),
complement(AL03538.1:32331..32562),
complement(AL03538.1:29831..31707),
complement(AL03538.1:29856),
complement(AL03538.1:28922..29051),
complement(AL03538.1:24293..24400),
complement(AL03538.1:17058..22895))
/locus_tag="RP3-477H23.1-001"
join(complement(88018..88119),complement(37948..38226),
complement(AL023281.1:96000..96147),
complement(AL023281.1:87084..87356),
complement(AL023281.1:85702..85832),
complement(AL035397.4:30093..30600),
complement(AL050313.6:101..1143),
complement(AL035453.4:18624..31904),
complement(AL035453.4:29823..30346),
complement(AL035453.4:25715..25824),
complement(AL035453.4:23459..23588),
complement(AL035453.4:20734..20952),
complement(AL035453.4:18624..18789),
complement(AL03538.1:69270..69410),
complement(AL03538.1:53292..53436),
complement(AL03538.1:40418..40597),
complement(AL03538.1:37584..38304),
complement(AL03538.1:35179..35303),
complement(AL03538.1:32331..32562),
complement(AL03538.1:29831..31707),
complement(AL03538.1:29856),
complement(AL03538.1:28922..29051),
complement(AL03538.1:24293..24400),
complement(AL03538.1:21265..22895))
/locus_tag="RP3-477H23.1-001"
/standard_name="OTTHUMP0000028696"
/codon_start=1
/protein_id="CAH73824.1"
/db_xref="GI:55660890"
/db_xref="InterPro:IPR001440"
/db_xref="InterPro:IPR011595"
/db_xref="InterPro:IPR011716"
/db_xref="InterPro:IPR011717"
/db_xref="UniProtKB/TrEMBL:Q5W189"
/translaton="MEQSPPPAPEPTQPTPARRRRRRPPSPASAPIPLFGADTTIG
QRSPGVGLVKEFVKVQSNQACHDGFHTATVLYNEALAVDPQCILYSNNR3AAY
MKIQYDKALDDAIKARLLNPKPKAYFRQGVALLQGRHADALAAFAAGLAQDPKSL
QLLVGVWEAAKMSPMRDSLEPTYQQLQKMKLDKSPFVVVSVVQGELLTAGHHGASVVV
```

```
misc_feature
43142..43210
/note="Single clone region. pUC subcloned to give 3x
coverage over region."
52224..52271
/note="Weak data"
69442..69505
/note="Weak data"
join(complement(AL117330.6:3554..26647),
complement(96139..120413))
/locus_tag="RP11-436C9.1-001"
join(complement(AL117330.6:25928..26647),
complement(AL117330.6:18562..18886),
complement(AL117330.6:9402..9526),
complement(AL117330.6:9136..9283),
complement(AL117330.6:3554..3644),
complement(120305..120413),complement(118402..118455),
complement(111901..111962),complement(108234..108333),
complement(105297..105383),complement(104269),
complement(103523..103638),complement(102428..102513),
complement(97531..97611),complement(96293..96382))
/locus_tag="RP11-436C9.1-001"
join(complement(AL117330.6:18562..18880),
complement(AL117330.6:9402..9526),
complement(AL117330.6:9136..9283),
complement(AL117330.6:3554..3644),
complement(120305..120413),complement(118402..118455),
complement(111901..111962),complement(108234..108333),
complement(105297..105383),complement(104269),
complement(103523..103638),complement(102428..102513),
complement(97531..97611),complement(96293..96382))
/locus_tag="RP11-436C9.1-001"
/standard_name="OTTHUMP0000028871"
/codon_start=1
```

```

/protein_id="CAH73823.1"
/db_xref="GI:5560889"

Query Match      91.6%; Score 17.4; DB 5; Length 124337;
Best Local Similarity 94.7%; Pred. No. 3.4e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19
    |||||
Db 123733 CTTTACTTCATAGTCTTTG 123715

RESULT 43
AC011354/c
LOCUS      133042 bp DNA linear PRI 27-JUN-2001
DEFINITION Homo sapiens chromosome 5 clone CTC-354F19, complete sequence.
AC011354
ACCESSION  AC011354.4 GI:14572125
VERSION    HTG.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homiidae; Homo.
REFERENCE  1 (bases 1 to 133042)
            DOE Joint Genome Institute and Stanford Human Genome Center.
            TITLE Direct Submission
            JOURNAL Unpublished
REFERENCE  2 (bases 1 to 133042)
            DOE Joint Genome Institute.
            TITLE Direct Submission
            JOURNAL Submitted (27-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
REFERENCE  3 (bases 1 to 133042)
            Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE     Direct Submission
JOURNAL   Submitted (27-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT   On Jun 27, 2001 this sequence version replaced gi:13699555.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.9% of Sequence;
            Estimated Total Number of Errors is 0.1.
            Location/Qualifiers
                1..133042
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
                /chromosome="5"
                /clones="CTC-354F19"

FEATURES             source
     source           1..133042
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
                        /chromosome="5"
                        /clones="CTC-354F19"

ORIGIN
Query Match      91.6%; Score 17.4; DB 5; Length 133042;
Best Local Similarity 94.7%; Pred. No. 3.5e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19
    |||||
Db 95584 CTTTACTTCATAGTCTTTG 95566

RESULT 44
AC011334/c
LOCUS      134413 bp DNA linear PRI 27-NOV-2001
DEFINITION Homo sapiens chromosome 5 clone CTC-235N17, complete sequence.
AC011334
ACCESSION  AC011334.5 GI:17105283
VERSION    HTG.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homiidae; Homo.
REFERENCE  1 (bases 1 to 134413)
            DOE Joint Genome Institute and Stanford Human Genome Center.
            TITLE Direct Submission
            JOURNAL Unpublished
REFERENCE  2 (bases 1 to 134413)
            DOE Joint Genome Institute.
            TITLE Direct Submission
            JOURNAL Submitted (06-OCT-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE  3 (bases 1 to 134413)
            Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE     Direct Submission
JOURNAL   Submitted (27-NOV-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT   On Nov 27, 2001 this sequence version replaced gi:12830127.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 100% of Sequence;
            Estimated Total Number of Errors is 0.
            Location/Qualifiers
                1..134413
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
                /chromosome="5"
                /clones="CTC-235N17"

ORIGIN
Query Match      91.6%; Score 17.4; DB 5; Length 134413;
Best Local Similarity 94.7%; Pred. No. 3.5e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19
    |||||
Db 22550 CTTTACTTCATAGTCTTTG 22532

RESULT 45
AP006464/c
LOCUS      141529 bp DNA linear PLN 28-JUL-2004
DEFINITION Oryza sativa (Japonica cultivar-group) genomic DNA, chromosome 9,
            BAC clone: B1054C11.
ACCESSION  AP006464
VERSION    AP006464.3 GI:50726670
KEYWORDS   Oryza sativa (japonica cultivar-group)
            Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP
            clade; Ehrhartoideae; Oryzaceae; Oryza.
ORIGIN
Reference        1
Sasaki, T., Matsumoto, T. and Katayose, Y.
Oryza sativa nipponbare (GA3) genomic DNA, chromosome 9, BAC
clone: B1054C11
Published Only in Database (2003)
2 (bases 1 to 141529)
Sasaki, T., Matsumoto, T. and Katayose, Y.
Direct Submission
Reference        2
Sasaki, T., Matsumoto, T. and Katayose, Y.
Submitted (30-MAY-2003) Takuji Sasaki, National Institute of
Agrobiological Sciences, Rice Genome Research Program; Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail: tsasaki@nias.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/,
Tel: 81-298-38-7441, Fax: 81-298-38-7468)
On Jul 27, 2004 this sequence version replaced gi:45752574.
Genes were predicted from the integrated results of the following:
GENSCAN (http://CCR-081.mit.edu/GENSCAN.html), FGENESH
(http://www.softberry.com/), GeneMark.hmm
(http://opal.biology.gatech.edu/GeneMark/), GlimmerM
(http://www.tigr.org/tdb/glimmer/glmr_form.html), RiceHMM

```

(<http://rgp.dna.affrc.go.jp/RiceHMM/>), SplicePredictor (<http://bioinformatics.tastate.edu/cgi-bin/sp.cgi>), sim4 (<http://globin.cse.psu.edu/html/docs/sim4.html>), gap2 (<http://www.tigr.org/software/glimmer/>), BLASTN and BLASTX. The genomic sequence was searched against NCBI NonRedundant Protein database, nr (<ftp://ncbi.nlm.nih.gov/blast/db>) and the cDNA sequence database at RGP or DDBJ. Protein homologies of the coding regions were searched against NCBI NonRedundant Protein database with BLASTP. ESTs represent the identified cDNA sequences using BLASTN with the corresponding DDBJ accession no. and RGP clone ID. Full-length cDNAs represent the identified cDNA sequences using BLASTN with the corresponding DDBJ accession no.

A gene with identity or significant homology to a protein is classified based on the protein name to indicate the homology level such as same name, 'putative', and '-like protein'. A gene without significant homology to any protein but with full-length cDNA or EST homology (covering almost the entire length of partial sequence) is classified as an 'unknown' protein. A gene predicted by two or more gene prediction programs is classified as a 'hypothetical' protein according to IRGSP standard. A gene predicted by a single gene prediction program is also classified as a probable 'hypothetical' protein and is included as a miscellaneous feature of the sequence.

The orientation of the sequence is from -21M13 to M13rev of the BAC clone. This sequence of B1054C11 clone has an overlap with B1106B03 (DDBJ: AP006449) clone at 5' end and an overlap with B1043F11 (DDBJ: AP006156) clone at 3' end. Detailed information on overlap and assembly quality together with annotation of this entry is available at <http://rgp.dna.affrc.go.jp/Genomeseq.html>

FEATURES	source	Location/Qualifiers
		1..141529
		/organism="Oryza sativa (japonica cultivar-group)"
		/mol_type="genomic DNA"
		/cultivar="Nipponbare"
		/db_xref="taxon:39947"
		/chromosome="9"
		/clone="B1054C11"
gene		complement(779..1069)
		/gene="B1054C11.1"
misc_feature		complement(779..1069)
		/gene="B1054C11.1"
		/note="hypothetical ORF predicted by GlimmerM
gene		this category is not included in IRGSP standard"
		4054..6208
		/gene="B1054C11.2"
misc_feature		4054..6208
		/gene="B1054C11.2"
		/note="probably inactive due to including frameshift(s) in CDS
		probably inactive due to including stop codon(s) in CDS
		pseudogene, putative reverse transcriptase"
		/pseudo
gene		7971..8195
mRNA		/gene="B1054C11.3"
		<7971..>8195
		/gene="B1054C11.3"
		/note="start and end point are not identified"
CDS		7971..8195
		/gene="B1054C11.3"
		/note="predicted by GlimmerM etc."
		/codon_start=1
		/product="hypothetical protein"
		/protein_id="BAD34388.1"
		/db_xref="GI:50726671"
		/translation="MAKLQDTLRMQBQQQAYEAAKQAKHTQOFYAQPQVVPFYAAQ
		APVYFAKPAQVAVRLTPQVTSVPRTAP"
gene		8808..10491
		/gene="B1054C11.4"
misc_feature		8808..10491
		/gene="B1054C11.4"
		/note="probably inactive due to 5'exon missing in CDS
		probably inactive due to including stop codon(s) in CDS

	pseudogene, RIRE2 GAG-POL precursor orf1"
/pseudo	
gene	10576..11713 /gene="B1054C11.5"
misc_feature	10576..11713 /gene="B1054C11.5"  note="probably inactive due to 3'exon missing in CDS probably inactive due to including frameshift(s) in CDS probably inactive due to including stop codon(s) in CDS pseudogene, RIRE2 GAG-POL precursor orf2" /pseudo
gene	complement(join(12879..12958,13076..13343,17753..17915, 18075..18227,18368..18485,18555..18642)) /gene="B1054C11.6"
misc_feature	complement(join(12879..12958,13076..13343,17753..17915, 18075..18227,18368..18485,18555..18642)) /gene="B1054C11.6"  note="hypothetical ORF predicted by GENSCAN this category is not included in IRGSP standard" 20255..20710 /gene="B1054C11.9"
repeat_unit	20255..20710 /gene="B1054C11.9"  note="5' Terminal repeat" rpt_type=Terminal 22332..22823 /gene="B1054C11.7" 22332..22823 /gene="B1054C11.7" /note="polyprotein-like" 23549..24028 /gene="B1054C11.8" 23549..24028 /gene="B1054C11.8" /note="pol polyprotein-like" 24296..28412 /gene="B1054C11.9"
misc_feature	24296..27636 /gene="B1054C11.9"  note="probably inactive due to including frameshift(s) in CDS pseudogene, RIRE2 GAG-POL precursor orf2" /pseudo
repeat_unit	27958..28412 /gene="B1054C11.9"  note="3' Terminal repeat" rpt_type=Terminal complement(join(30838..30884,31676..31753,32773..32922, 36455..36500) /gene="B1054C11.10" complement(join(30838..30884,31676..31753,32773..32922, 36455..36500) /gene="B1054C11.10"  note="hypothetical ORF predicted by GENSCAN this category is not included in IRGSP standard" complement(36818..37609) /gene="B1054C11.11" complement(36818..37609) /gene="B1054C11.11" /note="probably inactive due to including frameshift(s) in CDS probably inactive due to including stop codon(s) in CDS pseudogene, putative polypeptidein" /pseudo
gene	complement(join(39930..39959,40065..40202)) /gene="B1054C11.12"
misc_feature	complement(join(39930..39959,40065..40202)) /gene="B1054C11.12"  note="hypothetical ORF predicted by GlmmRM this category is not included in IRGSP standard"

```

gene      join(42044..42502,42615..43076)
           /gene="B1054C11.13"
misc_feature      join(42044..42502,42615..43076)
           /gene="B1054C11.13"
           /notes="transposase-like protein"
gene      43456..43906
           /gene="B1054C11.14"
mRNA      join(43456..43534,43614..43906)
           /gene="B1054C11.14"
           /notes="start and end point are not identified"
CDS       join(43456..43534,43614..43906)
           /gene="B1054C11.14"
           /notes="predicted by GeneMark.hmm etc."
           /codon_start=1
           /product="hypothetical protein"
           /protein_id="BAD34389.1"
           /db_xref="GI:50726672"
           /translation="MLAKAAQKNPKGHPHTKKHGSPIEKRRASWNTALEKILVEL
LHBNHTPEYRGQNGWTSIAWNKIVKFHFEDRYVCLTKSQIQEKELKREYRMLKEA
RKSGASWNNQRICIIAEPAI"
45057..47758
           /gene="B1054C11.15"
mRNA      join(45057..45138,47163..47758)
           /gene="B1054C11.15"

Query Match      91.6%  Score 17.4;  DB 4;  Length 141529;
Best Local Similarity 94.7%;  Pred. No. 3.5e+02;
Matches 18;  Conservative 0;  Mismatches 1;  Indels 0;  Gaps 0;

QY  1  CTTACTTCATAGTCCTTG 19
      |||||
Db  58588 CTTACTTCATAGTCCTCG 58570

RESULT 46
AC091155  AC091155 144430 bp DNA linear PRI 24-APR-2002
LOCUS      Homo sapiens chromosome 17, clone CTD1-2306D4, complete sequence.
DEFINITION AC091155
ACCESSION  AC091155
VERSION    AC091155.5 GI:20128191
KEYWORDS   HTG.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominiidae; Homo.
REFERENCE  1  (bases 1 to 144430)
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
TITLE      Barna,N., Bastien,V., Boguslavsky,L., Bouckghalter,B., Brown,A.,
JOURNAL    Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,
REFERENCE  Collins,S., Collumore,A., Cooke,P., DeArellano,K., Dewar,K.,
AUTHORS    Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D.,
JOURNAL    Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
REFERENCE  Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,
AUTHORS    Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRocque,K.,
TITLE      Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,
JOURNAL    MacLean,C., Macdonald,P., McKernan,K., Meldrim,J., Meneus,L.,
AUTHORS    Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,
TITLE      Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
JOURNAL    Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
AUTHORS    Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
TITLE      Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
JOURNAL    Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,H.,
AUTHORS    Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
TITLE      Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission

AC091155  AC091155 144430 bp DNA linear PRI 24-APR-2002
LOCUS      Homo sapiens chromosome 17, clone CTD1-2306D4, complete sequence.
DEFINITION AC091155
ACCESSION  AC091155
VERSION    AC091155.5 GI:20128191
KEYWORDS   HTG.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominiidae; Homo.
REFERENCE  1  (bases 1 to 144430)
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
TITLE      Barna,N., Bastien,V., Boguslavsky,L., Bouckghalter,B., Brown,A.,
JOURNAL    Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,
REFERENCE  Collins,S., Collumore,A., Cooke,P., DeArellano,K., Dewar,K.,
AUTHORS    Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D.,
JOURNAL    Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
REFERENCE  Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,
AUTHORS    Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRocque,K.,
TITLE      Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,
JOURNAL    MacLean,C., Macdonald,P., McKernan,K., Meldrim,J., Meneus,L.,
AUTHORS    Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,
TITLE      Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
JOURNAL    Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
AUTHORS    Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
TITLE      Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,
JOURNAL    Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,H.,
AUTHORS    Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
TITLE      Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission

```

# JOURNAL REFERENCE AUTHORS

Submitted (01-APR-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 144430)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Bouckghalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collumore,A., Cooke,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,H., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

## Direct Submission

# TITLE JOURNAL REFERENCE AUTHORS

Submitted (10-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

4 (bases 1 to 144430)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Bouckghalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collumore,A., Cooke,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,H., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

## Direct Submission

# TITLE JOURNAL COMMENT

Submitted (24-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Apr 10, 2002 this sequence version replaced gi:19683224.

All repeats were identified using RepeatMasker:

Smit, A.P.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

## Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L13231

Center clone name: 2306\_D\_4

## Location/Qualifiers

1. 144430

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="17"

# FEATURES

## source

# TITLE





This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate  
chemistry, or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by sequence  
from more than one subclone; and the assembly was confirmed by  
restriction digest.

## MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D.  
McPherson, Department of Genetics, Washington University, St. Louis  
MO. For additional information about the map position of this  
sequence, see <http://genome.wustl.edu/gsc>

## SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male  
donor, as described by Osoegawa, K., Woon, P. Y., Zhao, B., Frengen, E.,  
Tateno, M., Catanese, J. J., and de Jong, P. J. (1998) An improved  
approach for construction of bacterial artificial chromosome  
libraries. *Genomics* 51:1-8. The clone may be obtained either from  
Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong  
and coworkers at the Roswell Park Cancer Institute  
(<http://bacpac.med.buffalo.edu>)

VECTOR: pBACe3.6

## NEIGHBORING SEQUENCE INFORMATION:

Actual start of this clone is at base position 1 of RP11-724M22;  
actual end is at base position 150013 of RP11-724M22.

## FEATURES

source	1..150013
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="4"
	/map="4"
	/clone="RP11-724M22"
	/clone_lib="RPCI-11"
repeat_region	1604..1629
	/rpt_family="(T)n"
repeat_region	1940..2102
	/rpt_family="MER2_type"
repeat_region	3157..3344
	/rpt_family="MIR"
repeat_region	4358..4685
	/rpt_family="MER2_type"
repeat_region	4691..4928
	/rpt_family="MER1_type"
repeat_region	4938..5019
	/rpt_family="MIR"
repeat_region	5020..5323
	/rpt_family="Alu"
repeat_region	5324..5414
	/rpt_family="MIR"
repeat_region	5843..6691
	/rpt_family="L2"
repeat_region	6876..6913
	/rpt_family="AT_rich"
repeat_region	6949..6996
	/rpt_family="(TAGA)n"
repeat_region	7820..8324
	/rpt_family="L2"
repeat_region	8321..8405
	/rpt_family="MIR"
repeat_region	8600..8628
	/rpt_family="(TC)n"
misc_feature	9468..9733
	/note="similar to EST BE841757 (NID:g10274135)"
repeat_region	9666..10139
	/rpt_family="MaLR"
misc_feature	9780..10467
	/note="similar to EST AV731443 (NID:g10840864)"
repeat_region	11206..11282
	/rpt_family="Alu"

repeat_region	11375..11770
	/rpt_family="L2"
repeat_region	11789..11956
	/rpt_family="MER1_type"
repeat_region	12024..12298
	/rpt_family="L2"
repeat_region	12299..12612
	/rpt_family="Alu"
repeat_region	12400..12434
	/rpt_family="AT_rich"
repeat_region	12582..12622
	/rpt_family="(GAAA)n"
repeat_region	12813..12839
	/rpt_family="L2"
repeat_region	12885..13093
	/rpt_family="MIR"
repeat_region	13626..13755
	/rpt_family="MIR"
repeat_region	14039..14306
	/rpt_family="Alu"
repeat_region	14307..14716
	/rpt_family="ERVL"
repeat_region	16531..16825
	/rpt_family="Alu"
repeat_region	16922..17123
	/rpt_family="MIR"
repeat_region	18268..18649
	/rpt_family="MER2_type"
repeat_region	18650..18762
	/rpt_family="Alu"
repeat_region	18763..18980
	/rpt_family="MER2_type"
repeat_region	20151..20274
	/rpt_family="MIR"
misc_feature	20616..20947
	/note="similar to EST AI418542 (NID:g4264473) tg37a05.x1"
misc_feature	21519..21975
	/note="similar to EST N70371 (NID:g12226951) z444f11.s1"
repeat_region	21749..22169
	/note="similar to EST W05012 (NID:g1277732) za44f11.r1"
repeat_region	21988..22167
	/rpt_family="MIR"
repeat_region	22528..22753
	/rpt_family="MIR"
repeat_region	23375..23507
	/rpt_family="MIR"
repeat_region	23705..23800
	/rpt_family="MIR"
repeat_region	23794..23840
	/rpt_family="(CA)n"
repeat_region	23843..24138
	/rpt_family="Alu"
repeat_region	24148..24176
	/rpt_family="AT_rich"
repeat_region	24160..25313
	/rpt_family="L1"
repeat_region	26631..26655
	/rpt_family="(T)n"
repeat_region	27027..27049
	/rpt_family="(A)n"
repeat_region	30559..30724
	/rpt_family="MIR"
repeat_region	32770..33273
	/rpt_family="MaLR"
repeat_region	33431..33723
	/rpt_family="Alu"
repeat_region	33919..34025
	/rpt_family="CRI"
repeat_region	34049..34213
	/rpt_family="MIR"
repeat_region	34480..34782
	/rpt_family="Alu"
repeat_region	34762..34782

```

/rpt family=" (A)n"
34789. 34936
/rpt family="MIR"
35145. 35269
/rpt family="MIR"
35521. 35828
/rpt family="Alu"
35807. 35828
/rpt family=" (A)n"
36448. 36486

```

Query Match 91.6%; Score 17.4; DB 5; Length 150013;

Best Local Similarity 94.7%; Pred. No. 3.5e+02;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATGCTTTG 19

Db 97641 CATTACTTCATGCTTTG 97623

#### RESULT 48

AP003856  
LOCUS AP003856 158479 bp DNA linear PLN 05-OCT-2004  
DEFINITION Oriza sativa (japonica cultivar-group) genomic DNA, chromosome 1,  
PAC clone: P0697D09.  
ACCESSION AP003856 BA000010  
VERSION AP003856.4 GI:20161655

SOURCE Oriza sativa (japonica cultivar-group)

ORGANISM Oriza sativa (japonica cultivar-group)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP  
clade; Ehrhartoideae; Oryzaceae; Oryza.

#### REFERENCE

1  
AUTHORS Sasaki, T., Matsumoto, T., Yamamoto, K., Sakata, K., Baba, T.,  
Katayose, Y., Wu, J., Niimura, Y., Cheng, Z., Nagamura, Y.,  
Antonio, B. A., Kanamori, H., Hosokawa, S., Masukawa, M., Arikawa, K.,  
Chiden, Y., Hayaishi, M., Okamoto, M., Ando, T., Aoki, H., Arita, K.,  
Hamada, M., Harada, C., Hijishita, S., Honda, M., Ichikawa, Y., Ito, Y.,  
Itonuma, A., Iijima, M., Ikeda, M., Ikeno, M., Ito, S., Ito, T., Ito, Y.,  
Ito, Y., Iwabuchi, A., Kamiya, K., Karasawa, W., Katagiri, S.,  
Kikuta, A., Kobayashi, N., Kono, I., Machita, K., Maehara, T.,  
Mizuno, H., Mizubayashi, T., Mukai, Y., Nagasaki, H., Nakashima, M.,  
Nakama, Y., Nakamichi, Y., Nakamura, M., Namiki, N., Negishi, M.,  
Ohta, I., Ono, N., Saji, S., Sakai, K., Shibata, M., Shimokawa, T.,  
Shomura, A., Song, J., Takazaki, Y., Terasawa, K., Tsuji, K., Waki, K.,  
Yamagata, H., Yamane, H., Yoshiki, S., Yoshihara, R., Yukawa, K.,  
Zhong, H., Iwama, H., Endo, T., Ito, H., Hahn, J. H., Kim, H. I., Eun, M. Y.,  
Yano, M., Jiang, J. and Gojobori, T.  
The genome sequence and structure of rice chromosome 1  
Nature 420 (6913), 312-316 (2002)

2 (bases 1 to 158479)

12447438

Sasaki, T., Matsumoto, T. and Yamamoto, K.

Direct Submission

Submitted (05-JUL-2001) Takuji Sasaki, National Institute of

Agrobiological Sciences, Rice Genome Research Program; Kannondai

2-1-2, Tsukuba, Ibaraki 305-8602, Japan

(E-mail: tsusakiana@affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/,

Tel: 81-298-38-7441, Fax: 81-298-38-7468)

On Apr 16, 2002 this sequence version replaced gi:16904699.

Genes were predicted from the integrated results of the following:

GENSCAN (http://ccr-081.mit.edu/GENSCAN.html), GENESH

(http://www.softberry.com/), GeneMark.hmm

(http://opal.biology.gatech.edu/GeneMark/), Glimmer

(http://www.tigr.org/tdb/glimmer/glimr form.html), RiceHMM

(http://rgp.dna.affrc.go.jp/RiceHMM/), SplicePredictor

(http://bioinformatics.iastate.edu/cgi-bin/sp.cgi), sim4

(http://globin.cse.psu.edu/html/docs/sim4.html), gap2

(http://www.tigr.org/software/glimmer/), BLASTN and BLASTX. The

genomic sequence was searched against NCBI NonRedundant Protein

database, nr (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA

sequence database at RGP or DDBJ. Protein homologies of the coding

regions were searched against NCBI NonRedundant Protein database with BLASTP. ESTs represent the identified cDNA sequences using BLASTN with the corresponding DDBJ accession no. and RGP clone ID. Full-length cDNAs represent the identified cDNA sequences using BLASTN with the corresponding DDBJ accession no.

A gene with identity or significant homology to a protein is classified based on the protein name to indicate the homology level such as same name, 'putative-' and '-like protein'. A gene without significant homology to any protein but with full-length cDNA or EST homology (covering almost the entire length of partial sequence) is classified as an 'unknown' protein. A gene predicted by two or more gene prediction programs is classified as a 'hypothetical' protein according to IRGSP standard. A gene predicted by a single gene prediction program is also classified as a probable 'hypothetical' protein and is included as a miscellaneous feature of the sequence.

The orientation of the sequence is from SP6 to T7 of the PAC clone. This sequence of P0697D09 clone has an overlap with P0503E05 (DDBJ: AP003021) clone at 5' end. Detailed information on overlap and assembly quality together with annotation of this entry is available at <http://rgp.dna.affrc.go.jp/GenomesSeq.html>.

#### FEATURES

##### source

1. 158479

/organism="Oryza sativa (japonica cultivar-group)"

/mol\_type="genomic DNA"

/cultivar="Nipponbare"

/db\_xref="taxon:39947"

/chromosome="1"

/clones="P0697D09"

1754. 1984

/gene="P0697D09.1"

1754. 1984

/gene="P0697D09.1"

/note="hypothetical ORF

predicted by GlimmerM

this category is not included in IRGSP standard"

complement(2282..2515)

/gene="P0697D09.2"

complement(2282..2515)

/gene="P0697D09.2"

/note="hypothetical ORF

predicted by GlimmerM

this category is not included in IRGSP standard"

complement(2588..3167)

/gene="P0697D09.3"

complement(join(<2588..2913,3032..>3167))

/gene="P0697D09.3"

/note="start and end point are not identified"

complement(join(2588..2913,3032..3167))

/gene="P0697D09.3"

/note="predicted by GeneMark.hmm etc."

/codon\_start=1

/product="hypothetical protein"

/protein\_id="BAD53954.1"

/db\_xref="GI:53793491"

/translation="MERGCAETELAAAKPAVEAAHQSGSGGGGGLWLEATATER

GALAGKSSPESGARRGESAGLGNKGNDEGSTGVVYMGAGEGDKQGRQGGN

GRDGVATWARGGKPAIPAGRVGEKVEEWATTARSQVRRRRQF"

complement(3399..3725)

/gene="P0697D09.4"

complement(<3399..>3725)

/gene="P0697D09.4"

/note="start and end point are not identified"

complement(3399..3725)

/gene="P0697D09.4"

/note="predicted by GeneMark.hmm etc."

/codon\_start=1

/product="hypothetical protein"

/protein\_id="BAD53955.1"

/db\_xref="GI:53793492"

/translation="MGVHGPKYRRGRRAAANDSGARPQDGHRRHRWRSARGERR

KERRRGGAHQARHQREARRRRSSGHGNGPVPFREREPAQDGLGNKEGKLG

REEAHE"

##### gene

##### mRNA

##### CDS

```
gene
misc_feature
  join(4678..4782,5254..5361)
  /genes="P0697D09.5"
  join(4678..4782,5254..5361)
  /genes="P0697D09.5"
  /notes="hypothetical ORF
  predicted by GeneMark.hmm
  this category is not included in IRGSP standard"
  7882..8256
  /genes="P0697D09.6"
  <7882..>8256
  /genes="P0697D09.6"
  /note="start and end point are not identified"
  7882..8256
  /genes="P0697D09.6"
  /notes="predicted by GlimmerM etc."
  /codon_start=1
  /product="hypothetical protein"
  /protein_id="BAD53956.1"
  /db_xref="GI:53793493"
  /translation="MBEGDAGRRRWPWRCSPAHDAEVPAESDGRGVDGVVTRTTAN
  ATRLGAVASGRSGGGGKGEGDGAAMEGTGDTKRRERGGGPLYRPAGVDRGW
  DGRDFAGVGGRGGEEREAGFKN"
  complement(join(8318..12103,17938..18448))
  /genes="P0697D09.7"
  /complement(8318..8835)
  /genes="P0697D09.7"
  /notes="3' terminal repeat"
  /rpt_type="terminal
  complement(8863..12103)
  /genes="P0697D09.7"
  /notes="probably inactive due to including stop codon(s) in
  CDS
  pseudogene, polyprotein"
  /pseudo
  14751..15035
  /genes="P0697D09.8"
  14751..15035
  /genes="P0697D09.8"
  /notes="hypothetical ORF
  predicted by GlimmerM
  this category is not included in IRGSP standard"
  16900..17385
  /genes="P0697D09.9"
  <16900..>17385
  /genes="P0697D09.9"
  /notes="start and end point are not identified"
  16900..17385
  /genes="P0697D09.9"
  /notes="similar to Oryza sativa chromosome 1, OJ1460_H08.6"
  /codon_start=1
  /product="hypothetical protein"
  /protein_id="BAD53957.1"
  /db_xref="GI:53793494"
  /translation="MAVTRVAVRGOGPTGEGGAPRGTDVVTWTRGGHWHABG
  CRTDGIPIPRVDRTRAAKAAAGAHRLKPAPRWTAALPRCGGRRROPTASGGGA
  SAKMLAAAKWLGGGCGAMRYRGAAADGGRRRAAKWLGGVATKRRGRERNGR
  G"
  complement(17938..18448)
  /genes="P0697D09.7"
  /notes="5' terminal repeat"
  /rpt_type="terminal
  join(19302..19350,19637..19840,20866..21059)
  /genes="P0697D09.10"
  join(19302..19350,19637..19840,20866..21059)
  /genes="P0697D09.10"
  /notes="hypothetical ORF
  predicted by GENSCAN
  this category is not included in IRGSP standard"
  complement(24099..25801)
  /genes="P0697D09.11"
  complement(24099..25801)
  /genes="P0697D09.11"
  /note="probably inactive due to 5' exon missing in CDS
```

```
gene
  pseudogene, wall-associated kinase 5"
  /pseudo
  join(27260..27262,29203..29946)
  Query Match 91.6%; Score 17.4; DB 4; Length 158479;
  Best Local Similarity 94.7%; Pred No. 3.5e+02;
  Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
  QY - 1 CTTTACTTCATAGCTTTG 19
  ||||| ||||| ||||| |||||
  Db 105161 CTTTGCTTCATAGCTTTG 105179

RESULT 49
AP003556/c
LOCUS
DEFINITION
  AP003556 167007 bp DNA linear HTG 02-MAY-2001
  SEQUENCE, 28 unordered pieces.
ACCESSION
  AP003556.1 GI:13928405
VERSION
  HTG; HTGS PHASE1; HTGS DRAFT.
KEYWORDS
  Homo sapiens (human)
SOURCE
  ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
  Homnidae; Homo.
  1 (bases 1 to 167007)
  Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
  Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  Direct Submission
  Submitted (27-APR-2001) Masahira Hattori, The Institute of Physical
  and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
  1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
  (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
  Tel:81-45-503-9111, Fax:81-45-503-9170)
  ----- Genome Center
  Center: RIKEN Genomic Sciences Center (GSC)
  Center code: RIKEN
  Web site: http://hgp.gsc.riken.go.jp/
  Contact: hattori@gsc.riken.go.jp
  ----- Project Information
  Center project name: HumDraft11
  Center clone name: RP11-644G15
  ----- Summary Statistics
  Sequencing vector: PCR products; 100% of reads
  Chemistry: Dye-terminator ET-amersham; 100% of reads
  Assembly program: Phrap; version 0.990329
  Consensus quality: 161481 bases at least Q40
  Consensus quality: 163250 bases at least Q30
  Consensus quality: 163985 bases at least Q20
  Insert size: 164307; sum-of-contigs
  Quality coverage: 8.64x in Q20 bases; sum-of-contigs
  -----
  NOTE: This is a 'working draft' sequence. It currently consists of
  28 contigs. The true order of the pieces is not known and their
  order in this sequence record is arbitrary. Gaps between the
  contigs are represented as runs N, but the exact sizes of the gaps
  are unknown. This record will be updated with the finished sequence
  as soon as it is available and the accession number will be
  preserved
  1 16343 contig of 16343 bp in length
  16444 33154 contig of 16711 bp in length
  33255 45637 contig of 12383 bp in length
  45738 58174 contig of 12437 bp in length
  58275 69718 contig of 11444 bp in length
  69819 77379 contig of 7561 bp in length
  77480 84668 contig of 7189 bp in length
  84769 92027 contig of 7259 bp in length
  92128 99269 contig of 7142 bp in length
  99370 104879 contig of 5510 bp in length
  104980 111041 contig of 6062 bp in length
  111142 116604 contig of 5463 bp in length
  116705 122298 contig of 5594 bp in length
```

122399 126480 contig of 4082 bp in length  
126581 130116 contig of 3536 bp in length  
130217 133305 contig of 3089 bp in length  
133406 138289 contig of 4884 bp in length  
138390 141974 contig of 3585 bp in length  
142075 145634 contig of 3560 bp in length  
145735 148493 contig of 2759 bp in length  
148594 151947 contig of 3354 bp in length  
152048 155192 contig of 3145 bp in length  
155293 157139 contig of 1847 bp in length  
157240 159731 contig of 2492 bp in length  
159832 162489 contig of 2658 bp in length  
162590 164577 contig of 1988 bp in length  
164678 166329 contig of 1652 bp in length  
166430 167007 contig of 578 bp in length.

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 28 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 16343: contig of 16343 bp in length  
\* 16344 16443: gap of 100 bp  
\* 16444 33154: contig of 16711 bp in length  
\* 33155 33254: gap of 100 bp  
\* 33255 45637: contig of 12383 bp in length  
\* 45638 45737: gap of 100 bp  
\* 45738 58174: contig of 12437 bp in length  
\* 58175 58274: gap of 100 bp  
\* 58275 69718: contig of 11444 bp in length  
\* 69719 69818: gap of 100 bp  
\* 69819 77379: contig of 7561 bp in length  
\* 77380 77479: gap of 100 bp  
\* 77480 84668: contig of 7189 bp in length  
\* 84669 84769: gap of 100 bp  
\* 84769 92027: contig of 7259 bp in length  
\* 92028 92127: gap of 100 bp  
\* 92128 99269: contig of 7142 bp in length  
\* 99270 99369: gap of 100 bp  
\* 99370 104879: contig of 5510 bp in length  
\* 104880 104979: gap of 100 bp  
\* 104980 111041: contig of 6062 bp in length  
\* 111042 111141: gap of 100 bp  
\* 111142 116604: contig of 5463 bp in length  
\* 116605 116705: gap of 100 bp  
\* 116705 122299: contig of 5594 bp in length  
\* 122299 122399: gap of 100 bp  
\* 122399 126480: contig of 4082 bp in length  
\* 126481 126581: gap of 100 bp  
\* 126581 130217: contig of 3536 bp in length  
\* 130217 130216: gap of 100 bp  
\* 130217 133305: contig of 3089 bp in length  
\* 133306 133406: gap of 100 bp  
\* 133406 138289: contig of 4884 bp in length  
\* 138290 138390: gap of 100 bp  
\* 138390 141974: contig of 3585 bp in length  
\* 141975 142074: gap of 100 bp  
\* 142075 145634: contig of 3560 bp in length  
\* 145635 145734: gap of 100 bp  
\* 145735 148493: contig of 2759 bp in length  
\* 148494 148594: gap of 100 bp  
\* 148594 151947: contig of 3354 bp in length  
\* 151948 152047: gap of 100 bp  
\* 152048 155192: contig of 3145 bp in length  
\* 155193 155293: gap of 100 bp  
\* 155293 157139: contig of 1847 bp in length  
\* 157140 157240: gap of 100 bp  
\* 157240 159731: contig of 2492 bp in length  
\* 159732 159831: gap of 100 bp  
\* 159832 162489: contig of 2658 bp in length  
\* 162490 162589: gap of 100 bp

\* 162590 164577: contig of 1988 bp in length  
\* 164578 164677: gap of 100 bp  
\* 164678 166329: contig of 1652 bp in length  
\* 166330 166429: gap of 100 bp  
\* 166430 167007: contig of 578 bp in length.

FEATURES  
    source  
        1. 167007  
            /organism="Homo sapiens"  
            /mol\_type="genomic DNA"  
            /db\_xref="taxon:9606"  
            /chromosome="11"  
            /map="11q"  
            /clone="RP11-644G15"  
        1. 16343  
            /notes="assembly\_fragment"  
        16444..33154  
            /notes="assembly\_fragment"  
        33255..45637  
            /notes="assembly\_fragment"  
        45738..58174  
            /notes="assembly\_fragment"  
        58275..69718  
            /notes="assembly\_fragment"  
        69819..77379  
            /notes="assembly\_fragment"  
        77480..84668  
            /notes="assembly\_fragment"  
        84769..92027  
            /notes="assembly\_fragment"  
        92128..99269  
            /notes="assembly\_fragment"  
        99370..104879  
            /notes="assembly\_fragment"  
        104980..111041  
            /notes="assembly\_fragment"  
        111142..116604  
            /notes="assembly\_fragment"  
        116705..122298  
            /notes="assembly\_fragment"  
        122399..126480  
            /notes="assembly\_fragment"  
        126581..130116  
            /notes="assembly\_fragment"  
        130217..133305  
            /notes="assembly\_fragment clone\_end:SP6 vector\_side:left"  
        133406..138289  
            /notes="assembly\_fragment"  
        138390..141974  
            /notes="assembly\_fragment"  
        142075..145634  
            /notes="assembly\_fragment"  
        145735..148493  
            /notes="assembly\_fragment"  
        148594..151947  
            /notes="assembly\_fragment"  
        152048..155192  
            /notes="assembly\_fragment"  
        155293..157139  
            /notes="assembly\_fragment"  
        157240..159731  
            /notes="assembly\_fragment"  
        159832..162489  
            /notes="assembly\_fragment"  
        162590..164577  
            /notes="assembly\_fragment"  
        164678..166329  
            /notes="assembly\_fragment"  
        166430..167007  
            /notes="assembly\_fragment clone\_end:17 vector\_side:left"

Query Match 91.6%; Score 17.4; DB 12; Length 167007;  
Best Local Similarity 94.7%; Pred. No. 3.5e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

Oy      1 CTTTACTTCATAGTCTTTG 19
Db      23295 CTTTACTTCATGGTCTTTG 23277

RESULT 50
AC173916/c
LOCUS   AC173916.1 167341 bp DNA linear HTG 07-DEC-2005
DEFINITION
Aterlix albiiventris clone LB4-10706, WORKING DRAFT SEQUENCE, 11
unordered pieces.
ACCESSION
AC173916
VERSION
AC173916.1 GI:83131777
KEYWORDS
HTG; HTGS PHASE1; HTGS DRAFT.
SOURCE
Aterlix albiiventris
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Insectivora; Erinaceidae;
Erinaceinae; Aterlix.
REFERENCE
1 (bases 1 to 167341)
Antonelias,A., Ayele,K., Bass,D., Benjamin,B., Bera,J.,
Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S., Chu,G.,
Coleman,H., Franks,S., Fuksenko,T., Gestole,M., Greene,A., Guan,X.,
Gupta,J., Gurson,N., Haghighi,P., Han,E., Han,J., Hansen,N.,
Ho,S.-L., Hu,P., Hunter,G., Hurie,B., Idol,J.R., Kwong,P.,
Laric,P., Larson,S., Lee-Lin,S.-O., Legaspi,R., Madden,M.,
Maduro,Q.L., Maduro,V.B., Margulies,E.H., Mastello,C., Maskeri,B.,
McDowell,J., Montemayor,C., Mullikin,J.C., Park,M., Portnoy,M.E.,
Prasad,A., Puri,O., Rantz,K., Reddix-Dugue,N., Sante,A.,
Schandler,K., Schueler,M.G., Sison,C., Stantripop,S., Taye,A.,
Thomas,J.W., Thomas,P.J., Tsipouri,V., Ung,L., Vogt,J.L.,
Wetherby,K.D., Withers,T.R., Young,A. and Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 167341)
Green,E.D.
Direct Submission
Submitted (07-DEC-2005) NIH Intramural Sequencing Center, 5625
Fishers Lane, Rockville, MD 20852, USA
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc.zoo@nhgri.nih.gov
----- Project Information
Center project name: lyf
Center clone name: 107006
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 158069 bases at least Q40
Consensus quality: 160024 bases at least Q30
Consensus quality: 161532 bases at least Q20
Insert size: 165000; agarose-fp
Insert size: 166341; sum-of-contents
Quality coverage: 7.56x in Q20 bases; agarose-fp
Quality coverage: 7.50x in Q20 bases; sum-of-contents
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 2871: contig of 2871 bp in length
* 2872 2971: gap of unknown length
* 2972 5156: contig of 2185 bp in length
* 5157 5256: gap of unknown length
* 5257 8163: contig of 2907 bp in length
* 8164 8263: gap of unknown length

* 8264 12517: contig of 4254 bp in length
* 12518 12617: gap of unknown length
* 12618 19985: contig of 7368 bp in length
* 19986 25452: gap of unknown length
* 25453 25552: contig of 5367 bp in length
* 25553 31075: gap of unknown length
* 31076 31175: contig of 5523 bp in length
* 31176 55107: gap of unknown length
* 55108 55207: contig of 23932 bp in length
* 55208 77292: gap of unknown length
* 77293 102159: contig of 22085 bp in length
* 102160 102259: gap of unknown length
* 102260 167341: contig of 24767 bp in length
* 167342 167341: contig of 5082 bp in length.
FEATURES
Location/Qualifiers
1..167341
/organism="Aterlix albiiventris"
/mol_type="genomic DNA"
/db_xref="taxon:9368"
/clone="LB4-10706"
/clone_lib="LB4"
/notes="BAC resource: http://bacpac.chori.org/"
1..2871
/notes="assembly_fragment"
2872..2971
/estimated_length=unknown
2972..5156
/notes="assembly_fragment"
5157..5256
/estimated_length=unknown
5257..8163
/notes="assembly_fragment"
8164..8263
/estimated_length=unknown
8264..12517
/notes="assembly_fragment"
12518..12617
/vector_side:right"
12618..19985
/estimated_length=unknown
19986..20085
/notes="assembly_fragment"
20086..25452
/estimated_length=unknown
25453..25552
/notes="assembly_fragment"
25553..31075
/estimated_length=unknown
31076..31175
/notes="assembly_fragment"
31176..55107
/estimated_length=unknown
55108..55207
/notes="assembly_fragment"
55208..77292
/estimated_length=unknown
77293..77392
/estimated_length=unknown
77393..102159
/notes="assembly_fragment"
102160..102259
/estimated_length=unknown
102260..167341
/notes="assembly_fragment"
vector_side:right"

ORIGIN
Query Match 91.6%; Score 17.4; DB 12; Length 167341;
Best Local Similarity 94.7%; Pred. No. 3.5e+02;

```

Matches	18;	Conservative	0;	Mismatches	1;	Indels	0;	Gaps	0;
Oy	1	CTTTACTTCATAGTCTTTG	19						
Db	144700	CTTTACTTCATAGTCTTTG	144682						
RESULT 51									
AP003021									
LOCUS	AP003021	168258 bp	DNA	linear	PLN 05-OCT-2004				
DEFINITION	Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1,								
ACCESSION	PAC clone: P0503E05.								
VERSION	AP003021 BA000010								
KEYWORDS	AP003021.2 GI:13486765								
SOURCE	Oryza sativa (japonica cultivar-group)								
ORGANISM	Oryza sativa (japonica cultivar-group)								
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP clade; Ehrhartoideae; Oryzeae; Oryza.								
AUTHORS	1 Sasaki, T., Matsumoto, T., Yamamoto, K., Sakata, K., Baba, T., Katayose, Y., Wu, J., Niimura, Y., Cheng, Z., Nagamura, Y., Antonio, B.A., Kanamori, H., Hosokawa, S., Masukawa, M., Arikawa, K., Chiden, Y., Hayashi, M., Okamoto, M., Ando, T., Aoki, H., Arita, K., Hamada, M., Harada, C., Hijishita, S., Honda, M., Ichikawa, Y., Idonuma, A., Iijima, M., Ikeda, M., Ikeno, M., Ito, S., Ito, T., Ito, Y., Ito, Y., Iwabuchi, A., Kamiya, K., Karasawa, W., Katagiri, S., Kikuta, A., Kobayashi, N., Kono, I., Machita, K., Maehara, T., Mizuno, H., Mizubayashi, T., Mukai, Y., Nagasaki, H., Nakashima, M., Nakama, Y., Nakamichi, Y., Nakamura, M., Namiki, N., Negishi, M., Ohta, I., Ono, N., Sai, S., Sakai, K., Shibata, M., Shimokawa, T., Shomura, A., Song, J., Takazaki, Y., Terasawa, K., Tsuji, K., Waki, K., Yamagata, H., Yamane, H., Yoshiki, S., Yoshihara, R., Yukawa, K., Zhong, H., Iwama, H., Endo, T., Ito, H., Hahn, J.H., Kim, H.I., Eun, M.Y., Yano, M., Jiang, J. and Gojobori, T.								
TITLE	The genome sequence and structure of rice chromosome 1								
JOURNAL	Nature 420 (6913), 312-316 (2002)								
PUBLISHED	12447438								
REFERENCE	2 (bases 1 to 168258)								
AUTHORS	Sasaki, T., Matsumoto, T. and Yamamoto, K.								
TITLE	Direct Submission								
JOURNAL	Submitted (06-DEC-2000) Takuji Sasaki, National Institute of Agrobiological Sciences, Rice Genome Research Program; Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan								
COMMENT	(E-mail: tsasaki@nias.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/, Tel: 81-298-38-7441, Fax: 81-298-38-7468) On Mar 28, 2001 this sequence version replaced gi:11602828. Genes were predicted from the integrated results of the following: GENSCAN (http://CCR-081.mit.edu/GENSCAN.html), FGENESH (http://opal.biology.gatech.edu/Genemark/), GlimmerH (http://www.tigr.org/tdb/glimmer/glmr.form.html), RiceHMM (http://rgp.dna.affrc.go.jp/RiceHMM/), SlicerPredictor (http://bioinformatics.iastate.edu/cgi-bin/sp.cgi), sim4 (http://globin.cse.psu.edu/html/docs/sim4.html), gap2 (http://www.tigr.org/software/glimmer/), BLASTN and BLASTX. The genomic sequence was searched against NCBI Nonredundant Protein database, nr (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at RGP or DDBJ. Protein homologues of the coding regions were searched against NCBI Nonredundant Protein database with BLASTP. ESTs represent the identified cDNA sequences using BLASTN with the corresponding DDBJ accession no. and RGP clone ID. Full-length cDNAs represent the identified cDNA sequences using BLASTN with the corresponding DDBJ accession no. A gene with identity or significant homology to a protein is classified based on the protein name to indicate the homology level such as same name, 'putative-' and '-like protein'. A gene without significant homology to any protein but with full-length cDNA or EST homology (covering almost the entire length of partial sequence) is classified as an 'unknown' protein. A gene predicted by two or more gene prediction programs is classified as a 'hypothetical' protein according to IRGSP standard. A gene								

FEATURES	Location/Qualifiers
source	1..168258
	/organism="Oryza sativa (japonica cultivar-group)"
	/mol_type="genomic DNA"
	/cultivar="Nipponbare"
	/db_xref="taxon:39947"
	/chromosome="1"
gene	join(1659..1716,1912..2030)
	/gene="P0503E05.1"
misc_feature	join(1659..1716,1912..2030)
	/gene="P0503E05.1"
	/note="hypothetical ORF predicted by GENSCAN
gene	this category is not included in IRGSP standard"
	complement(2316..4423)
	/gene="P0503E05.2"
misc_feature	complement(2316..4423)
	/gene="P0503E05.2"
	/note="contains full-length cDNA(s): AK071457 non-coding transcript probably inactive due to including stop codon(s) in CDS"
gene	complement(4443..5185)
	/gene="P0503E05.3"
mrna	complement(4443..5185)
	/gene="P0503E05.3"
	/note="supported by full-length cDNA(s): AK107733"
CDS	complement(4541..4924)
	/gene="P0503E05.3"
	/note="contains full-length cDNA(s): AK107733"
	/codon_start=1
	/product="unknown protein"
	/protein_id="BAD52626.1"
	/db_xref="GI:53791504"
	/translation="MSDSLPRPAEKRRRRSPRLLPMWCCACVRRRRRCKKHAGAP ASAPALVPASAPAPASRCRCRCRKAPACAVEAGGVAGGLERAPHRDSRGG RGEREREYDRWDHVEGILYRQKLK"
gene	join(5101..5174,5333..5411,5604..5702)
mrna	join(<5101..5174,5333..5411,5604..5702)
	/gene="P0503E05.4"
	/gene="P0503E05.4"
	/note="start and end point are not identified"
CDS	join(5101..5174,5333..5411,5604..5702)
	/gene="P0503E05.4"
	/note="predicted by FGENESH etc."
	/codon_start=1
	/product="hypothetical protein"
	/protein_id="BAD52627.1"
	/db_xref="GI:53791505"
	/translation="MGDGGEASLEPKRLDEIDSIWNSRNVGRGSMKYEAQVAFGLLF FSKQPFNDIWDQAFQKYTVDFVQPLEHNDVGVTSRVVF"
gene	7045..7404
mrna	<7045..>7404
	/gene="P0503E05.5"
	<7045..>7404
	/note="start and end point are not identified"
CDS	7045..7404
	/gene="P0503E05.5"
	/note="predicted by FGENESH etc."
	/codon_start=1
	/product="hypothetical protein"
	/protein_id="BAD52628.1"
	/db_xref="GI:53791506"
	/translation="MARKDBAEGLKARLSLPKSKRCGDKDGGGGRKKTKTYLVNLGV

predicted by a single gene prediction program is also classified as a probable 'hypothetical' protein and is included as a miscellaneous feature of the sequence.  
The orientation of the sequence is from T7 to SP6 of the PAC clone. This sequence of P0503E05 clone has an overlap with P0460H02 (DDBJ: AP003257) clone at 5' end and with P0697D09 (DDBJ: AP003856) clone at 3' end. Detailed information on overlap and assembly quality together with annotation of this entry is available at <http://rgp.dna.affrc.go.jp/GenomeSeq.html>.

```
CCSKEGVEULMWLPAPAMERLVAEVARGLVGVDAHARDRWGGRREHMYRGALIGCP
IILEIFFSTFKMPHQAA"
/gene .8260
/misc_feature
8012. .8260
/gene="P0503E05.6"
/gene="P0503E05.6"
/notes="hypothetical ORF
predicted by GlimmerM
this category is not included in IRGSP standard"
8660. .9384
/gene="P0503E05.7"
/mRNA <8660. .9384
/gene="P0503E05.7"
/misc_feature
8660. .9384
/notes="supported by full-length cDNA (s) : AK064749"
/gene="P0503E05.7"
/notes="contains full-length cDNA(s) : AK064749
non-coding transcript
probably inactive due to including stop codon(s) in CDS"
11036. .11562
/gene="P0503E05.8"
/mRNA join(11036. .11237,11531. .>11562)
/gene="P0503E05.8"
/notes="start and end point are not identified"
11036. .11237,11531. .11562
/gene="P0503E05.8"
/notes="predicted by FGENSEH etc."
/codon_start=1
/product="hypothetical protein"
/protein_id="BA052629.1"
/db_xref="GI:53791507"
MSALVRPGQPDPTPSLIAGVIRGEIPHFVHAT"
/gene="P0503E05.9"
/mRNA join(12016. .12506,15527. .15967)
/gene="P0503E05.9"
/notes="supported by full-length cDNA (s) : AK060356"
12074. .12506,15527. .15977)
/gene="P0503E05.9"
/notes="supported by full-length cDNA (s) : AK059622"
12190. .12506,15527. .15957)
/gene="P0503E05.9"
/notes="contains EST (s) : D47200(S12386), AU096192 (S12386)
contains full-length cDNA(s) : AK060356, AK059622
similar to Arabidopsis thaliana, AAM67044"
/codon_start=1
/product="unknown protein"
/protein_id="BAB40002.1"
/db_xref="GI:13486769"
/translators="MESSRGKPGKLGSGGGAADFYSRRGYVTGAGAAPPLAAGSR
APPVDPCCVLRVFLVLTASAVVMAADRSQTTVQIAAGEELAPLRVPTAKWTYS
SAFYFVYANVAFPSAALAAVRRRSVVPVMDLVAMALLFSVAGAAOFGLLG
ERGNHVRAKVCVDPGPFPCERAAVVALIAFADLVLLMLTLTIHKASSY"
17745. .20869
/gene="P0503E05.10-2"
/misc_feature 17745. .20869
/gene="P0503E05.10-2"
/notes="contains full-length cDNA (s) : AK071463
non-coding transcript
probably inactive due to including stop codon(s) in CDS"
17745. .20846
/gene="P0503E05.10-3"

Query Match 91.6%; Score 17.4; DB 4; Length 168258;
Best Local Similarity 94.7%; Pred. No. 3.5e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGCTTTG 19
||||| ||||| ||||| |||||
Db 156379 CTTTGCTTCATAGCTTTG 156397
```

```
RESULT 52
AC015570
LOCUS
DEFINITION
AC015570.5 GI:7329242
HTG; HTGS_PHASE1; HTGS_DRAFT.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 168509)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 18, clone RP11-453F17
Unpublished
2 (bases 1 to 168509)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Perreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., MacDonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 26, 2000 this sequence version replaced gi:6778593.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L1024
Center clone name: 453 F.17
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 161651 bases at least Q40
Consensus quality: 164748 bases at least Q30
Consensus quality: 166260 bases at least Q20
Insert size: 168000; agarose-fp
Insert size: 167609; sum-of-contigs
Quality coverage: 5.2 in Q20 bases; agarose-fp
Quality coverage: 5.2 in Q20 bases; sum-of-contigs
-----
```

```
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 4392: contig of 4392 bp in length
* 4393 4492: gap of 100 bp
* 4493 8988: contig of 4496 bp in length
* 8989 9088: gap of 100 bp
* 9089 18712: contig of 9624 bp in length
* 18713 18812: gap of 100 bp
```





```

* 3882 3981: gap of 100 bp
* 3882 6990: contig of 3009 bp in length
* 6991 7090: gap of 100 bp
* 7091 11767: contig of 4677 bp in length
* 11768 11867: gap of 100 bp
* 11868 16650: contig of 4783 bp in length
* 16651 16750: gap of 100 bp
* 16751 20824: contig of 4074 bp in length
* 20825 20924: gap of 100 bp
* 20925 26002: contig of 5078 bp in length
* 26003 26102: gap of 100 bp
* 26103 33403: contig of 7301 bp in length
* 33404 33503: gap of 100 bp
* 33504 40903: contig of 7400 bp in length
* 40904 41003: gap of 100 bp
* 41004 51483: contig of 10479 bp in length
* 51483 51582: gap of 100 bp
* 51583 66690: contig of 15108 bp in length
* 66691 66790: gap of 100 bp
* 66791 78793: contig of 12003 bp in length
* 78794 78893: gap of 100 bp
* 78894 97795: contig of 18903 bp in length
* 97797 97896: gap of 100 bp
* 97897 125633: contig of 27736 bp in length
* 125634 125733: gap of 100 bp
* 125733 168703: contig of 42971 bp in length.
FEATURES
    source
        1..168703
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="11"
            /map="11"
            /clone="RP11-136114"
            /clone_lib="RPCI-11 Human Male BAC"
        1..3881
            /note="assembly_fragment"
        3882..3981
            /estimated_length=100
        3982..6990
            /note="assembly_fragment"
        6991..7090
            /estimated_length=100
        7091..11767
            /note="assembly_fragment"
        11768..11867
            /estimated_length=100
        11868..16650
            /note="assembly_fragment"
        16651..16750
            /estimated_length=100
        16751..20824
            /note="assembly_fragment"
        20825..20924
            /estimated_length=100
        20925..26002
            /note="assembly_fragment"
        26003..26102
            /estimated_length=100
        26103..33403
            /note="assembly_fragment"
        33404..33503
            /estimated_length=100
        33504..40903
            /note="assembly_fragment"
        40904..41003
            /estimated_length=100
        41004..51482
            /note="assembly_fragment"
        51483..51582
            /estimated_length=100
        51583..66690
            /note="assembly_fragment"

```

```

gap
    66691..66790
        /estimated_length=100
    66791..78793
        /note="assembly_fragment"
    78794..78893
        /estimated_length=100
    78894..97796
        /note="assembly_fragment"
    97797..97896
        /estimated_length=100
    97897..125632
        /note="assembly_fragment"
    clone_end:SP6
    vector_side:right"
    125633..125732
        /estimated_length=100
    125733..168703
        /note="assembly_fragment"
    clone_end:I7
    vector_side:right"

ORIGIN
    Query Match          91.6%;   Score 17.4;   DB 12;   Length 168703;
    Best Local Similarity 94.7%;   Pred. No. 3.5e+02;
    Matches 18;   Conservative 0;   Mismatches 1;   Indels 0;   Gaps 0;

QY      1 CTTTACTTCATAGCTCTTTG 19
        |||||
Db       86430 CTTTACTTCATGGCTTTG 86412

RESULT 54
AC124725/c          168983 bp   DNA   linear   ROD 11-NOV-2003
LOCUS              Mus musculus BAC clone RP23-459F16 from chromosome 2, complete
DEFINITION
AC124725          AC124725.3   GI:25189024
VERSION
KEYWORDS          HTG.
SOURCE            Mus musculus (house mouse)
ORGANISM          Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidae; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 168983)
Harkins,R., Cotton,M., Mangiapanello,L. and Dignan,G.
The sequence of Mus musculus BAC clone RP23-459F16
JOURNAL
Unpublished (2001)
REFERENCE
2 (bases 1 to 168983)
Wilson,R.
Sequencing of Mus musculus
JOURNAL
Unpublished (2001)
REFERENCE
3 (bases 1 to 168983)
McPherson,J.D. and Waterston,R.H.
Direct Submission
Submitted (15-JUN-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE
4 (bases 1 to 168983)
McPherson,J.D. and Waterston,R.H.
Direct Submission
Submitted (30-OCT-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE
5 (bases 1 to 168983)
McPherson,J.D. and Waterston,R.H.
Direct Submission
Submitted (23-NOV-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE
6 (bases 1 to 168983)
Wilson,R.
Direct Submission
Submitted (11-NOV-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Nov 23, 2002 this sequence version replaced gi:24421750.
COMMENT

```

```

----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@wustl.edu
----- Summary Statistics
-----
Center project name: M_BA0459F16
-----

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

#### SOURCE INFORMATION:

The RPCI-23 BAC Library has been constructed by Kazutoyo Osegawa and Minako Tatenio in the laboratory of Pieter de Jong (<http://www.chori.org>) from female C57BL/6J mouse kidney and/or brain genomic DNA. The clone and detailed information can be obtained from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

#### NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone.

#### FEATURES

```

source
    1..168983
        /location=Qualifiers
        /organism="Mus musculus"
        /mol_type="genomic DNA"
        /db_xref="taxon:10090"
        /chromosome="2"
        /map="2"
        /clone="RP23-459F16"
        /clone_lib="RPCI-23"
        524..652
            /rpt_family="Alu"
        663..733
            /rpt_family="ID"
        755..943
            /rpt_family="B2"
        2009..2063
            /rpt_family="ERV1"
        2755..2908
            /rpt_family="L1"
        2908..3107
            /rpt_family="L1"
        4208..4300
            /rpt_family="MIR"
        4414..4503
            /rpt_family="Alu"
        4779..5170
            /rpt_family="L2"
        5558..5765
            /rpt_family="ERVK"
        6198..7098
            /rpt_family="L1"
        8020..8337
            /rpt_family="MaLR"
        8346..8444
            /rpt_family="B2"

repeat_region
    11192..11327
        /rpt_family="L1"
repeat_region
    11330..11636
        /rpt_family="L1"
repeat_region
    11706..11776
        /rpt_family="Alu"
repeat_region
    11788..11861
        /rpt_family="L1"
repeat_region
    11957..12216
        /rpt_family="B4"
repeat_region
    14630..15279
        /rpt_family="L1"
repeat_region
    15503..15853
        /rpt_family="ERVK"
repeat_region
    17183..17239
        /rpt_family="ERVK"
repeat_region
    17259..17425
        /rpt_family="B4"
repeat_region
    17512..17626
        /rpt_family="Alu"
repeat_region
    17673..17841
        /rpt_family="B2"
repeat_region
    17988..18189
        /rpt_family="B2"
repeat_region
    18998..19686
        /rpt_family="L1"
repeat_region
    19680..19758
        /rpt_family="L1"
repeat_region
    19963..20043
        /rpt_family="L1"
repeat_region
    20098..20334
        /rpt_family="B2"
repeat_region
    20358..20734
        /rpt_family="MaLR"
repeat_region
    20735..21572
        /rpt_family="MaLR"
repeat_region
    23136..23331
        /rpt_family="B2"
repeat_region
    23386..23889
        /rpt_family="L1"
repeat_region
    24016..24482
        /rpt_family="L1"
repeat_region
    24619..24939
        /rpt_family="MaLR"
repeat_region
    26223..26268
        /rpt_family="ERVL"
repeat_region
    26784..26877
        /rpt_family="L1"
repeat_region
    26906..28545
        /rpt_family="L1"
repeat_region
    28552..28739
        /rpt_family="L1"
repeat_region
    29793..29923
        /rpt_family="B4"
repeat_region
    34162..34240
        /rpt_family="MIR"
repeat_region
    34838..35122
        /rpt_family="MaLR"
repeat_region
    35215..35609
        /rpt_family="MaLR"
repeat_region
    36013..36200
        /rpt_family="MaLR"
repeat_region
    38402..38629
        /rpt_family="L1"
repeat_region
    38632..38884
        /rpt_family="MaLR"
repeat_region
    38885..40759
        /rpt_family="L1"
repeat_region
    40760..40923
        /rpt_family="B2"
repeat_region
    40986..41188
        /rpt_family="L1"
repeat_region
    41189..41226

```

```

/rpt_family="MaLR"
42017. .42125
/rpt_family="Alu"
42163. .42340
/rpt_family="L1"
42799. .43002
/rpt_family="B2"
43338. .43712
/rpt_family="MaLR"
43715. .44006
/rpt_family="MaLR"
43999. .44218
/rpt_family="MaLR"
44257. .44594

Query Match          91.6%; Score 17.4; DB 6; Length 168983;
Best Local Similarity 94.7%; Pred. No. 3.5e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCTTTG 19
|| ||||| ||||| |||||
Db 136616 CTCTACTTCATAGTCTTTG 136598

RESULT 55
AP003174
LOCUS Homo sapiens genomic DNA, chromosome 11q, clone:RP11-136I14,
DEFINITION
ACCESSION AP003174
VERSION AP003174.2 GI:17425259
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens genomic DNA
Published Only in Database (2001)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (09-FEB-2001) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsukumi-Ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
On Dec 7, 2001 this sequence version replaced gi:12862394.

FEATURES
source
1..170156
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="RP11-136I14"

ORIGIN
Query Match          91.6%; Score 17.4; DB 5; Length 170156;
Best Local Similarity 94.7%; Pred. No. 3.5e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCTTTG 19
|| ||||| ||||| |||||
Db 104965 CTTTACTTCATAGTCTTTG 104983

RESULT 56
AC127007
LOCUS Rattus norvegicus clone CH230-342K24, *** SEQUENCING IN PROGRESS
DEFINITION
ACCESSION AC127007
VERSION AC127007.3 GI:25139714
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 170956)
Muzny,D.,Marie., Metzker,M., Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Angulano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,B.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregiorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpach,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,S., Nair,L.,
Nankervis,C., Neal,D., Newton,S., Nguyen,N., Norris,S.,
Nwaokemele,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steinle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmari,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,X., Zhao,S., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 170956)
Worley,K.C.
Direct Submission
Submitted (12-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 170956)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (20-NOV-2002) Human Genome Sequencing Center, Department

```

of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 On Nov 20, 2002 this sequence version replaced gi:23196239.  
 The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: RAFO  
 Center clone name: CH230-342K24

----- Summary Statistics

Assembly program: Phrap; version 0.990329  
 Consensus quality: 158599 bases at least Q40  
 Consensus quality: 160475 bases at least Q30  
 Consensus quality: 161407 bases at least Q20  
 Estimated insert size: 161556; sum-of-contigs estimation  
 Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

-----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

\* 1 165449: contig of 165449 bp in length  
 \* 165450 165549: gap of unknown length  
 \* 165550 170956: contig of 5407 bp in length.

#### FEATURES

source  
 1. .170956  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-342K24"

misc\_feature  
 1. .1394  
 /note="wgs\_end\_extension  
 clone\_end:T7"

misc\_feature  
 4483. .5354  
 /note="clone\_boundary  
 clone\_end:T7"

misc\_feature  
 site:  
 end\_sequence:BZ209789"

72516. .94000  
 /note="clone\_boundary  
 clone\_end:Sp6"

misc\_feature  
 site:  
 end\_sequence:BZ209791"

163069. .165449  
 /note="wgs\_end\_extension  
 clone\_end:Sp6"

165450. .165549  
 /estimated\_length=unknown

#### ORIGIN

.Query Match 91.6%; Score 17.4; DB 12; Length 170956;  
 Best Local Similarity 94.7%; Pred. No. 3.5e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGCTTTG 19  
 Db 144157 CTTTATTTCATAGCTTTG 144175

#### RESULT 57

AC164008/c  
 LOCUS AC164008 171582 bp DNA linear ROD 10-AUG-2005  
 DEFINITION Mus musculus chromosome 1, clone RP23-311L17, complete sequence.  
 ACCESSION AC164008  
 VERSION AC164008.2 GI:72096358  
 KEYWORDS HTG.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 171582)

AUTHORS Birren,B., Nusbaum,C. and Lander,E.

TITLE Mus musculus chromosome 1, clone RP23-311L17

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 171582)

AUTHORS Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Collamore,A., Cook,A., Cooke,P., Corum,B., Dearellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R., MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (15-JUN-2005) Broad Institute of MIT and Harvard, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE 3 (bases 1 to 171582)

AUTHORS Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Collamore,A., Cook,A., Cooke,P., Corum,B., Dearellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R., MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (15-AUG-2005) Broad Institute of MIT and Harvard, 320 Charles Street, Cambridge, MA 02141, USA

On Aug 10, 2005 this sequence version replaced gi:67764029.

All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
----- Genome Center  
Center: Broad Institute of MIT and Harvard  
Center code: WfBR  
Web site: <http://www.seq.wi.mit.edu>  
Contact: [sequence\\_submissions@broad.mit.edu](mailto:sequence_submissions@broad.mit.edu)  
----- Project Information  
Center project name: J5706  
Center clone name: 311\_L\_17  
-----

Some of the sequence contained within base pairs 1 - 85171 was  
stolen from accession AC102786 [WICGR Project L19659]  
Some of the sequence contained within base pairs 107515 to the end  
of the clone was stolen from accession AC102703 [WICGR Project  
L19314].

## FEATURES

source	Location/Qualifiers
1. .171582	/organism="Mus musculus"
	/mol_type="genomic DNA"
	/db_xref="taxon:10090"
	/chromosome="1"
	/map="1"
	/clone="RP23-311L17"
	/clone_lib="RPCI-23 Female Mouse BAC"
	/complement(4. .49)
repeat_region	/rpt_family="L1_Mur2"
repeat_region	/complement(52. .305)
repeat_region	/rpt_family="L1_Mm"
repeat_region	/complement(301. .3064)
repeat_region	/rpt_family="L1VL1"
repeat_region	/complement(3033. .3500)
repeat_region	/rpt_family="L1Md F"
repeat_region	/complement(3480. .3641)
repeat_region	/rpt_family="L1Md F"
repeat_region	/complement(3702. .4004)
repeat_region	/rpt_family="L1VL1"
repeat_region	/complement(4048. .7454)
repeat_region	/rpt_family="L1VL4"
repeat_region	7701. .8661
repeat_region	/rpt_family="L1Md F"
repeat_region	8703. .12865
repeat_region	/rpt_family="L1Md F"
repeat_region	12867. .13866
repeat_region	/rpt_family="L1Md F2"
repeat_region	13980. .14004
repeat_region	/rpt_family="AT_rich"
repeat_region	14215. .14365
repeat_region	/rpt_family="B1_Mm"
repeat_region	14366. .14407
repeat_region	/rpt_family="(CAAA)n"
repeat_region	14501. .18604
repeat_region	/rpt_family="L1VL4"
repeat_region	18605. .19066
repeat_region	/rpt_family="IAPLTR2_Mm"
repeat_region	19067. .20864
repeat_region	/rpt_family="L1_Mus3"
repeat_region	20873. .21012
repeat_region	/rpt_family="B1_Mm"
repeat_region	21450. .21478
repeat_region	/rpt_family="AT_rich"
repeat_region	21755. .21778
repeat_region	/rpt_family="AT_rich"
repeat_region	22760. .23066
repeat_region	/rpt_family="Lx6"
repeat_region	23067. .23202
repeat_region	/rpt_family="B1_Mm"
repeat_region	23203. .23225
repeat_region	/rpt_family="(CAAA)n"
repeat_region	23226. .23836
repeat_region	/rpt_family="Lx6"

repeat_region	23862. .24156	/rpt_family="Lx6"
repeat_region	complement(25203. .25355)	/rpt_family="MER5A"
repeat_region	26206. .26246	/rpt_family="(TTTG)n"
repeat_region	26247. .26467	/rpt_family="L1_Mus3"
repeat_region	complement(26588. .26799)	/rpt_family="B4A"
repeat_region	27201. .27246	/rpt_family="AT_rich"
repeat_region	27278. .27334	/rpt_family="(TTTTG)n"
repeat_region	complement(27335. .27481)	/rpt_family="B1_Mus2"
repeat_region	complement(27571. .28185)	/rpt_family="L1_Mm"
repeat_region	28173. .28610	/rpt_family="Lx7"
repeat_region	28611. .28784	/rpt_family="B2_Mm2"
repeat_region	28785. .28940	/rpt_family="Lx7"
repeat_region	complement(28942. .29231)	/rpt_family="RMER4B"
repeat_region	29232. .29336	/rpt_family="(TAGA)n"
repeat_region	complement(29337. .29408)	/rpt_family="RMER4B"
repeat_region	29450. .29802	/rpt_family="Lx7"
repeat_region	29803. .30125	/rpt_family="ORR1A3"
repeat_region	30135. .30355	/rpt_family="Lx6"
repeat_region	30907. .30972	/rpt_family="(TAGG)n"
repeat_region	31000. .31025	/rpt_family="(TA)n"
repeat_region	31087. .31358	/rpt_family="RMER6C"
repeat_region	31359. .31734	/rpt_family="Lx2"
repeat_region	31735. .32124	/rpt_family="RMER6C"
repeat_region	complement(32375. .32557)	/rpt_family="Lx8"
repeat_region	complement(32827. .33407)	/rpt_family="Lx9"
repeat_region	complement(33411. .33843)	/rpt_family="(TA)n"
repeat_region	33894. .33917	/rpt_family="L1_Rod"
repeat_region	33918. .34227	/rpt_family="AT_rich"
repeat_region	35066. .35096	/rpt_family="Lx9"
repeat_region	35108. .35176	/rpt_family="Lx9"
repeat_region	35334. .35506	/rpt_family="MER2"

Query Match 91.6%; Score 17.4; DB 6; Length 171582;  
Best Local Similarity 94.7%; Pred.No. 3.5e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCCTTTG 19

Db 158637 CTTTCTTCATAGTCCTTTG 158619

RESULT 58

AC148231/c  
 LOCUS AC148231 176362 bp DNA linear HTG 17-FEB-2004  
 DEFINITION Didelphis virginiana clone LB3-44H21, WORKING DRAFT SEQUENCE, 7  
 ordered pieces.  
 ACCESSION AC148231 GI:42558291  
 VERSION HTG; HTGS PHASE2; HTGS DRAFT.  
 KEYWORDS Didelphis virginiana (North American opossum)  
 SOURCE Didelphis virginiana  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Metatheria; Didelphimorphia; Didelphidae; Didelphis.  
 REFERENCE 1 (bases 1 to 176362)  
 AUTHORS Cheng, J.-F., Hamilton, M., Peng, Y., Mukherjee, S., Hosseini, R.,  
 Peng, Z., Malinov, I. and Rubin, E.M.  
 TITLE Direct Submission  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 176362)  
 AUTHORS Cheng, J.-F., Hamilton, M., Peng, Y., Mukherjee, S., Hosseini, R.,  
 Peng, Z., Malinov, I. and Rubin, E.M.  
 TITLE Direct Submission  
 JOURNAL Submitted (17-FEB-2004) Genome Sciences, Lawrence Berkeley National  
 Laboratory, 1 Cyclotron Rd., Berkeley, CA 94720, USA

Sequence Produced by Berkeley PGA  
 Web site: <http://pga.lbl.gov>  
 Center Code: PGABERK  
 Center Project Name: O148  
 Bac Clone Name: LB3-44H21

This sequence has been compared to sequences of other species  
 using Vista (<http://www-gsd.lbl.gov/VISTA>). The results can be  
 viewed at:  
[http://pga.lbl.gov/cgi-bin/search\\_cvcgd?type=nt&value=LIPC](http://pga.lbl.gov/cgi-bin/search_cvcgd?type=nt&value=LIPC)

The order-orientation of the draft sequence was accomplished by  
 using:  
 Avid (<http://baboon.math.berkeley.edu/mavid/>),  
 Lagan (<http://lagan.stanford.edu/>) and paired end information.

Funding agent: Programs for Genomic Applications (NHUBI)

If the Bac Library Name is LBL to LBL4, please see website  
 for the description: <http://www-gsd.lbl.gov/cheng/BAC.html>  
 These libraries are available through the BACPAC Resources Center:  
<http://www.chori.org/bacpac/libraryres.htm> as LBNL-1 to LBNL-4.

#### Summary Statistics:

Sequencing vector: plasmid; pUC18  
 Chemistry: Dye-terminator Big Dye  
 Assembly program: Phrap version 0.950329.  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 7 contigs. Gaps between the contigs  
 \* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* provided by the submitter.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.  
 \* 1 1086: contig of 1086 bp in length  
 \* 1087 1186: gap of unknown length  
 \* 1187 17979: contig of 16793 bp in length  
 \* 17980 18079: gap of unknown length  
 \* 18080 60719: contig of 42640 bp in length  
 \* 60820 60819: gap of unknown length  
 \* 60820 73619: contig of 12800 bp in length  
 \* 73620 73719: gap of unknown length  
 \* 73720 87458: contig of 13739 bp in length  
 \* 87459 87558: gap of unknown length  
 \* 87559 166569: contig of 79011 bp in length  
 \* 166570 166669: gap of unknown length  
 \* 166670 176362: contig of 9693 bp in length.

#### FEATURES

Location/Qualifiers

source 1. 176362  
 /organism="Didelphis virginiana"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9267"  
 /clone="LB3-44H21"  
 1087..1186  
 /estimated\_length=unknown  
 17980..18079  
 /estimated\_length=unknown  
 60720..60819  
 /estimated\_length=unknown  
 73620..73719  
 /estimated\_length=unknown  
 87459..87558  
 /estimated\_length=unknown  
 166570..166669  
 /estimated\_length=unknown

#### ORIGIN

Query Match 91.6%; Score 17.4; DB 12; Length 176362;  
 Best Local Similarity 94.7%; Pred. No. 3.5e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 CTTTACTTCATAGTCTTTG 19  
 |||||

Db 71363 CTTTACTTCATAGACTTTG 71345

#### RESULT 59

AL845354 178468 bp DNA linear ROD 24-JUN-2003  
 LOCUS Mouse DNA sequence from clone RP23-318L10 on chromosome 2, complete  
 DEFINITION sequence.  
 ACCSSION AL845354  
 VERSION AL845354.13 GI:32187955  
 KEYWORDS HTG.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muroidea; Muridae; Murinae; Mus.  
 Tracey, A.  
 Direct Submission  
 Submitted (24-JUN-2003) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
[humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk) Clone requests: [clonerequest@sanger.ac.uk](mailto:clonerequest@sanger.ac.uk)  
 On Jun 24, 2003 this sequence version replaced GI:32131063.  
 Sequence from the Mouse Genome Sequencing Consortium whole genome  
 shotgun may have been used to confirm this sequence. Sequence data  
 from the whole genome shotgun alone has only been used where it has  
 a phred quality of at least 30.

----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: <http://www.sanger.ac.uk>  
 Contact: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)

#### REFERENCE

AUTHORS  
 TITLE  
 JOURNAL

#### COMMENT

During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >=  
 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest, except on the rare  
 occasion of the clone being a YAC.  
 The following abbreviations are used to associate primary accession

numbers given in the feature table with their source databases:  
 En: EMBL, Sw: SWISSPROT, Tr: TREMBL, Wp: WORMPEP; Information  
 on the WORMPEP database can be found at  
[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) RP23-318L10 is  
 from the RPCI-23 Mouse BAC Library  
 constructed by the group of Pieter de Jong.  
 For further details see <http://www.chori.org/bacpac/home.htm>  
 VECTOR: pBACe3.6.

FEATURES  
 source  
 1..178468  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /chromosome="2"  
 /clone="RP23-318L10"  
 /clone\_lib="RPCI-23"  
 ORIGIN  
 Query Match 91.6%; Score 17.4; DB 6; Length 178468;  
 Best Local Similarity 94.7%; Pred.No. 3.5e+02;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATGCTTTG 19  
 |||||  
 Db 138089 CTTTACTTCATGCTTTG 138107

RESULT 60  
 AL512635/c  
 LOCUS  
 DEFINITION  
 AL512635 178631 bp DNA linear PRI 18-MAY-2005  
 Human DNA sequence from clone RP11-15P13 on chromosome 9 Contains a  
 survival of motor neuron 1, telomeric (SMN1) pseudogene and the 3'  
 end of the MLLT3 gene for myeloid/lymphoid or mixed-lineage  
 leukemia (trithorax homolog, Drosophila); translocated to, 3 (AF9),  
 complete sequence.

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 AL512635.8 GI:16409123  
 HTG; AF9; MLLT3; SMN1.  
 Homo sapiens (human)  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Hominiidae; Homo.

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 Pearce,A.  
 Direct Submission  
 Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries: [vegas@sanger.ac.uk](mailto:vegas@sanger.ac.uk)  
 Clone requests: [clonerequest@sanger.ac.uk](mailto:clonerequest@sanger.ac.uk)  
 On Oct 24, 2001 this sequence version replaced gi:15131291.

COMMENT  
 The following abbreviations are used to associate primary accession  
 numbers given in the feature table with their source databases:

En: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information  
 on the WORMPEP database can be found at  
[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence  
 was generated from part of bacterial clone contigs of human  
 chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping  
 Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/Chr9>  
 RP11-15P13 is from the library RPCI-11.1 constructed by the group  
 of Pieter de Jong. For further details see  
<http://www.chori.org/bacpac/home.htm>  
 VECTOR: pBACe3.6

----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: <http://www.sanger.ac.uk>  
 Contact: [vegas@sanger.ac.uk](mailto:vegas@sanger.ac.uk)  
 -----  
 Draft Sequence Produced by Whitehead Institute/MIT Center for  
 Genome Research, 320 Charles Street,  
 Cambridge, MA 02141, USA  
<http://www-seq.wi.mit.edu>

This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >=  
 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one subclone; and the assembly was confirmed by restriction digest,  
 except on the rare occasion of the clone being a YAC.

FEATURES  
 Location/Qualifiers  
 1..178631  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="9"  
 /clone="RP11-15P13"  
 /clone\_lib="RPCI-11.1"  
 1  
 /note="Clone\_left\_end: RP11-15P13"  
 53156  
 /note="Clone\_right\_end: RP11-594J14"  
 148943..149774  
 /locus\_tag="RP11-15P13.1-001"  
 /pseudo  
 148943..149774  
 /locus\_tag="RP11-15P13.1-001"  
 /note="match: Proteins: AAC83178 AAH15308 O02771 O35876  
 P97801 Q16637 Q96J51"  
 /pseudo  
 /codon\_start=1  
 149506  
 /locus\_tag="RP11-15P13.1-001"  
 /note="Clone\_left\_end: RP11-336O12"  
 /pseudo  
 join(complement(AL354879.12:14212..14510),  
 complement(AL354879.12:12621..12801),  
 complement(AL513498.2:124696..124778),  
 complement(AL513498.2:116115..116258),  
 complement(AL513498.2:81713..82417),  
 complement(AL513498.2:33661..33736),  
 complement(AL513498.2:31468..31597),  
 complement(178239..178338),complement(172305..172376),  
 complement(171022..171093),complement(159162..164071))  
 /gene="MLLT3"  
 /locus\_tag="RP11-15P13.2-001"  
 join(complement(AL354879.12:14212..14510),  
 complement(AL354879.12:12621..12801),  
 complement(AL513498.2:124696..124778),  
 complement(AL513498.2:116115..116258),  
 complement(AL513498.2:81713..82417),  
 complement(AL513498.2:33661..33736),  
 complement(AL513498.2:31468..31597),  
 complement(178239..178338),complement(172305..172376),  
 complement(171022..171093),complement(159162..164071))  
 /gene="MLLT3"  
 /locus\_tag="RP11-15P13.2-001"  
 /product="myeloid/lymphoid or mixed-lineage leukemia  
 (trithorax homolog, Drosophila)"; translocated to, 3"  
 /note="match: ESTs: AA406302.1 AA42076.1 AA443284.1  
 AI336478.1 AI554049.1 AI620408.1 AI739108.1 AI768397.1  
 AL535503.1 AL565947.1 AW044186.1 AW163788.1 AW302520.1  
 AW451255.1 AW572468.1 BF104863.1 BF510928.1 BG940265.1  
 BG958983.1 BI550359.1 BI561777.1 BI710808.1 BM018568.1  
 BU623899.1 CA418678.1 HI503.1 H9306.1 N56884.1  
 match: cDNAs: BC036089.1 DI6688.1 LI3744.1"  
 complement(159162)  
 /gene="MLLT3"  
 /locus\_tag="RP11-15P13.2-001"  
 complement(159188..159193)  
 /gene="MLLT3"  
 /locus\_tag="RP11-15P13.2-001"  
 join(complement(AL513498.2:50373..50479),  
 complement(AL513498.2:33661..33736),  
 complement(AL513498.2:31468..31597),  
 complement(178239..178338),complement(172305..172376),  
 complement(159162..164071))  
 /polyA\_site  
 /locus\_tag="RP11-15P13.2-001"  
 complement(159193)  
 /polyA\_signal  
 /gene



complement (162574. .164071))  
 /gene="MLLT3"  
 /locus tag="RP11-15P13.2-006"  
 join(complement(ALS13498.2:50373. .50479),  
 complement(ALS13498.2:33661. .33736),  
 complement(ALS13498.2:31468. .31597),  
 complement(178239. .178338), complement(172305. .172376),  
 complement(162574. .164071))  
 /gene="MLLT3"  
 /locus tag="RP11-15P13.2-006"  
 /product="myeloid/lymphoid or mixed-lineage leukemia  
 (trithorax homolog, Drosophila)"; translocated to, 3"  
 /note="match: cDNAs: AK019458.1"  
 join(complement(ALS13498.2:31468. .31561),  
 complement(178239. .178338), complement(163790. .164071))  
 /gene="MLLT3"  
 /locus tag="RP11-15P13.2-007"  
 join(complement(ALS13498.2:31468. .31561),  
 complement(178239. .178338), complement(163790. .164071))  
 /gene="MLLT3"  
 /locus tag="RP11-15P13.2-007"  
 /product="myeloid/lymphoid or mixed-lineage leukemia  
 (trithorax homolog, Drosophila)"; translocated to, 3"  
 /note="match: ESTs: N66033.1"  
 join(complement(ALS13498.2:50373. .50463),  
 complement(172305. .172376), complement(171022. .171093),  
 complement(163850. .164071))  
 /gene="MLLT3"  
 /locus tag="RP11-15P13.2-005"  
 join(complement(ALS13498.2:50373. .50463),  
 complement(172305. .172376), complement(171022. .171093),  
 complement(163850. .164071))  
 /gene="MLLT3"  
 /locus tag="RP11-15P13.2-005"  
 /product="myeloid/lymphoid or mixed-lineage leukemia  
 (trithorax homolog, Drosophila)"; translocated to, 3"  
 /note="match: ESTs: BU664780.1"  
 join(complement(ALS13498.2:50373. .50461),  
 complement(ALS13498.2:33661. .33736),  
 complement(ALS13498.2:31468. .31597),  
 complement(178239. .178338), complement(172305. .172376),  
 complement(171022. .171093), complement(163903. .164071))  
 /gene="MLLT3"  
 /locus tag="RP11-15P13.2-002"  
 join(complement(ALS13498.2:50373. .50461),  
 complement(ALS13498.2:33661. .33736),  
 complement(ALS13498.2:31468. .31597),  
 complement(178239. .178338), complement(172305. .172376),  
 complement(171022. .171093), complement(163903. .164071))  
 /gene="MLLT3"  
 /locus tag="RP11-15P13.2-002"  
 /product="myeloid/lymphoid or mixed-lineage leukemia  
 (trithorax homolog, Drosophila)"; translocated to, 3"  
 /note="match: ESTs: BF243978.1 BG940264.1 BM793722.1"  
 join(complement(ALS13498.2:31468. .31580),  
 complement(178239. .178338), complement(172305. .172376),  
 complement(163940. .164071))  
 /gene="MLLT3"  
 /locus tag="RP11-15P13.2-006"  
 /standard name="OTTHUMP0000021129"  
 /codon start=1  
 /product="myeloid/lymphoid or mixed-lineage leukemia  
 (trithorax homolog, Drosophila)"; translocated to, 3"  
 /protein\_id="CAH70707.1"  
 /db\_xref="GI:55665053"  
 /db\_xref="UniProtKB/TrEMBL:Q5VV07"  
 /translations="MKDLHSDNEEDSEVEDNDSEMPVNRGGSRRVSLSDG  
 SDSESSASPLHHEPPPLTKTNQILVKSPIKQSKSDKIKNGEKDVIYNLIE  
 TGHFHINTFFDLCSLKTIVRKQSLYLETSGTS"  
 join(complement(AL354879.12:14212. .14223),  
 join(complement(AL354879.12:12621. .12801),  
 complement(ALS13498.2:124696. .124778),  
 complement(ALS13498.2:116115. .116258),

complement(ALS13498.2:81713. .82417),  
 complement(ALS13498.2:33661. .33736),  
 complement(ALS13498.2:31468. .31597),  
 complement(178239. .178338), complement(172305. .172376),  
 complement(171022. .171093), complement(163940. .164071))  
 /gene="MLLT3"  
 /locus tag="RP11-15P13.2-002"  
 /standard name="OTTHUMP0000021127"  
 /note="match: proteins: BAA04090 P43568 Q14768 Q8VDR6  
 Q99MK4"  
 Query Match 91.6%; Score 17.4; DB 5; Length 178631;  
 Best Local Similarity 94.7%; Pred. No. 3.5e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 CTTTACTTCATCTCTTTG 19  
 |||||  
 Db 66553 CTTTACTTCATCTCTTTG 66535  
 RESULT 61  
 AC011949  
 LOCUS 179112 bp DNA linear HTG 03-APR-2000  
 DEFINITION Homo sapiens clone RP11-17E12, WORKING DRAFT SEQUENCE, 10 unordered  
 pieces.  
 AC011949  
 AC011949.3 GI:7387331  
 HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 KEYWORDS Homo sapiens (human)  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 179112)  
 Birren, B., Linton, L., Nussbaum, C. and Lander, E.  
 Homo sapiens chromosome, clone RP11-17E12  
 Unpublished  
 REFERENCE 2 (bases 1 to 179112)  
 Birren, B., Linton, L., Nussbaum, C., Lander, E., Allen, N., Anderson, M.,  
 Baldwin, J., Barna, N., Beckerly, R., Boguslavsky, L., Boukhgaiter, B.,  
 Brown, A., Castelle, A., Colangelo, M., Collins, S., Collymore, A.,  
 Cooke, P., DeArelano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,  
 Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,  
 Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,  
 Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,  
 Lehoczy, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,  
 McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrim, J.,  
 Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,  
 Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,  
 Strange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
 Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,  
 Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (16-OCT-1999) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Apr 3, 2000 this sequence version replaced gi:6453950.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L3628  
 Center clone name: 17 E 12  
 ----- Summary Statistics  
 Sequencing vector: M13; M7815; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 150637 bases at least Q40  
 Consensus quality: 166570 bases at least Q30

Consensus quality: 173803 bases at least Q20  
Insert size: 182000; agarose-fp  
Insert size: 178212; sum-of-contigs  
Quality coverage: 4.0 in Q20 bases; agarose-fp  
Quality coverage: 4.1 in Q20 bases; sum-of-contigs

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 10 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence.  
\* as soon as it is available and the accession number will  
\* be preserved.

1 3511: contig of 3511 bp in length  
\* 3512: gap of 100 bp  
\* 3612: contig of 3092 bp in length  
\* 6703: gap of 100 bp  
\* 6704: contig of 5633 bp in length  
\* 12436: gap of 100 bp  
\* 12437: 12536: gap of 100 bp  
\* 12537: 18649: contig of 6113 bp in length  
\* 18650: 18749: gap of 100 bp  
\* 31179: contig of 12430 bp in length  
\* 31279: gap of 100 bp  
\* 44707: contig of 13428 bp in length  
\* 44807: gap of 100 bp  
\* 69716: contig of 24909 bp in length  
\* 69816: gap of 100 bp  
\* 99052: contig of 29236 bp in length  
\* 99152: gap of 100 bp  
\* 12097: contig of 32945 bp in length  
\* 132098: 132197: gap of 100 bp  
\* 132198: 179112: contig of 46915 bp in length.

## FEATURES

source  
1. .179112  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="RP11-17E12"  
/clone\_lib="RPCI-11 Human Male BAC"  
misc\_feature  
1. .3511  
/note="assembly\_fragment"  
gap  
3512. .3611  
/estimated\_length=100  
misc\_feature  
3612. .6703  
/note="assembly\_fragment"  
gap  
6704. .6803  
/estimated\_length=100  
misc\_feature  
6804. .12436  
/note="assembly\_fragment"  
gap  
12437. .12536  
/estimated\_length=100  
misc\_feature  
12537. .18649  
/note="assembly\_fragment"  
gap  
18650. .18749  
/estimated\_length=100  
misc\_feature  
18750. .31179  
/note="assembly\_fragment"  
gap  
31180. .31279  
/estimated\_length=100  
misc\_feature  
31280. .44707  
/note="assembly\_fragment"  
gap  
44708. .44807  
/estimated\_length=100  
misc\_feature  
44808. .69716  
/note="assembly\_fragment"  
clone\_end:17  
vector\_side:left  
69717. .69816  
/estimated\_length=100  
misc\_feature  
69817. .99052  
/note="assembly\_fragment"  
clone\_end:SP6

vector\_side:right"  
99053. .99152  
/estimated\_length=100  
misc\_feature  
99153. .132097  
/note="assembly\_fragment"  
gap  
132098. .132197  
/estimated\_length=100  
misc\_feature  
132198. .179112  
/note="assembly\_fragment"  
ORIGIN  
Query Match 91.6%; Score 17.4; DB 12; Length 179112;  
Best Local Similarity 94.7%; Pred. No. 3.5e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 CTTTACTTCATAGTCTTTG 19  
|||||  
Db 42157 CTTTACTTCATAGTCTTTG 42175  
|||||  
RESULT 62  
AC022743/c 182466 bp DNA linear PRI 01-MAY-2002  
LOCUS Homo sapiens chromosome 18, clone RP11-789C7, complete sequence.  
DEFINITION AC022743  
ACCESSION AC022743.7 GI:20148189  
VERSION HTG.  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 182466)  
Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
TITLE Homo sapiens chromosome 18, clone RP11-789C7  
JOURNAL Unpublished  
REFERENCE  
2 (bases 1 to 182466)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Bada,F.,  
Boguslavskiy,L., Bouckgalter,B., Brown,A., Burket,G., Castle,A.,  
Choepe,Y., Collangelo,M., Collins,S., Collymore,A., Cooke,P.,  
DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,  
Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,  
Gardyna,S., Grant,G., Hagos,B., Headford,A., Horton,L.,  
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,  
Macdonald,P., Marguis,N., McEwan,P., McGurk,A., McKernan,K.,  
McPheeters,R., Meldrum,J., Meneus,L., Morrow,J., Naylor,J.,  
Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,  
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,  
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,  
Zimmer,A. and Zody,M.  
TITLE Direct Submission  
JOURNAL Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
REFERENCE  
3 (bases 1 to 182466)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,  
Bouckgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,  
Chazaro,B., Choepe,Y., Colangelo,M., Collins,S., Collymore,A.,  
Cooke,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,  
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
Kamat,A., Karatas,A., Kells,C., Larocque,K., Lamazares,R.,  
Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,  
MacLean,C., Macdonald,P., Major,J., Marquis,N., Mathews,C.,  
McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Meneus,L.,  
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,  
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,  
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,

Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

# TITLE

JOURNAL  
 Submitted (14-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

## REFERENCE

## AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Collangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., DeAtellano, K., Dewar, K., Diaz, J. S., Dodge, S., Fero, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McSwan, P., McKernan, K., Meldrum, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

## Direct Submission

Submitted (01-MAY-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Apr 14, 2002 this sequence version replaced gi:18598798.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

## FEATURES

## source

Location/Qualifiers  
 1..182466  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="18"  
 /map="18"  
 /clone="RP11-789C7"  
 /clone\_lib="RPC1-11 Human Male BAC"  
 1130..1500  
 /rpt\_family="MLT11"  
 2328..2362  
 /rpt\_family="TA)n"  
 2850..3400  
 /rpt\_family="11PA5"  
 3401..3434  
 /rpt\_family="AT\_rich"  
 3530..3557  
 /rpt\_family="AT\_rich"  
 3585..3623  
 /rpt\_family="AT\_rich"  
 complement(4501..4989)  
 /rpt\_family="MER39"  
 5235..5260

repeat\_region  
 complement(5653..5707)  
 /rpt\_family="MLT2A1"  
 5708..5754  
 /rpt\_family="TG)n"  
 5776..5795  
 /rpt\_family="TA)n"  
 complement(5799..5851)  
 /rpt\_family="HERVL"  
 5852..6005  
 /rpt\_family="TA)n"  
 complement(6006..6392)  
 /rpt\_family="MLT2A1"  
 complement(6393..11940)  
 /rpt\_family="HERVL-A1"  
 complement(11941..11988)  
 /rpt\_family="MLT2A1"  
 complement(11976..12030)  
 /rpt\_family="LTR66"  
 12035..12125  
 /rpt\_family="TA)n"  
 complement(12126..12530)  
 /rpt\_family="MLT2A1"  
 14161..14302  
 /rpt\_family="L2"  
 complement(14339..14641)  
 /rpt\_family="AluSk"  
 14652..14754  
 /rpt\_family="L2"  
 15170..15393  
 /rpt\_family="MIR"  
 15430..15947  
 /rpt\_family="LTR53"  
 complement(17981..18304)  
 /rpt\_family="AluOo"  
 18553..18611  
 /rpt\_family="TA)n"  
 18648..18925  
 /rpt\_family="AluYa5"  
 complement(19028..19344)  
 /rpt\_family="L1ME1"  
 19345..19564  
 /rpt\_family="MER63A"  
 complement(19565..20399)  
 /rpt\_family="L1ME1"  
 21259..21299  
 /rpt\_family="AT\_rich"  
 21377..21447  
 /rpt\_family="AT\_rich"  
 23879..23993  
 /rpt\_family="MIR3"  
 24183..24203  
 /rpt\_family="AT\_rich"  
 24759..24919  
 /rpt\_family="MIR3"  
 25817..25852  
 /rpt\_family="CA)n"  
 complement(25931..26039)  
 /rpt\_family="MIR"  
 26380..26845  
 /rpt\_family="MER66A"  
 26929..26955  
 /rpt\_family="AT\_rich"  
 27057..27082  
 /rpt\_family="(A)n"  
 27340..27682  
 /rpt\_family="LTR16A"  
 28151..28222  
 /rpt\_family="CT-rich"  
 28262..28396  
 /rpt\_family="CT-rich"  
 complement(28763..28767)  
 /note="<30 qual SNGL region"

unsure

```

unsure      complement(31426..31658)
unsure      /notes="single clone coverage"
unsure      complement(31557..31561)

Query Match      91.6%; Score 17.4; DB 5; Length 182466;
Best Local Similarity 94.7%; Pred. No. 3.5e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19
    |||||
Db 43399 CTTTACTTCATAGTCTTTG 43381

RESULT 63
AC097585/c
LOCUS      182709 bp DNA linear HTG 01-MAR-2002
DEFINITION Sus scrofa clone RP44-35K6, WORKING DRAFT SEQUENCE, 5 ordered
            pieces.
ACCESSION  AC097585
VERSION    AC097585.2 GI:19033547
KEYWORDS   HTG; HTGS PHASE2; HTGS_DRAFT.
SOURCE     Sus scrofa (pig)
ORGANISM   Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae;
            Sus.
REFERENCE  1 (bases 1 to 182709)
            Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
            Blakeley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
            Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
            Ho,S.-L., Idol,J.R., Karlins,E., Laric,P., Lee-Lin,S.-Q.,
            Legaapi,R., Maduro,Q.L., Maduro,V.B., Masiello,C., Maskeri,B.,
            Mastrian,S.D., McCloskey,J.C., McDowell,J., Pearson,R., Prasad,A.,
            Stantripop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W.,
            Turgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
            Young,A., Zhang,L.-H. and Green,E.D.
            NISC Comparative Sequencing Initiative
            Unpublished
REFERENCE  2 (bases 1 to 182709)
            Green,E.D.
            Direct Submission
            Submitted (19-OCT-2001) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE  3 (bases 1 to 182709)
            Green,E.D.
            Direct Submission
            Submitted (01-MAR-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
            On Mar 1, 2002 this sequence version replaced gi:16271923.
            ----- Genome Center
            Center: NIH Intramural Sequencing Center
            Center code: NISC
            Web site: http://www.nisc.nih.gov
            Contact: nisc_mouse@hgri.nih.gov
            ----- Project Information
            Center project name: ctb
            Center clone name: 035K06

The sequence data in this record represents an 'enhanced'
version of a Phase 2 submission. Specifically, the indicated
order and orientation of each sequence contig has been
established using one or more of the following: read-pair
data from individual subclones, overlaps with neighboring
clones, alignment with available reference sequence (e.g.,
human), and/or confirmation by PCR testing. In addition,
the sequence assembly is based on at least 8X average
coverage in Q20 bases and has been reviewed to rule out
gross misassemblies, the low-quality ends of sequence
contigs have been trimmed away, and each base is associated
with a Phrap-derived quality score.
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads

```

Assembly program: Phrap; version 0.990319  
Consensus quality: 181870 bases at least Q40  
Consensus quality: 182166 bases at least Q30  
Consensus quality: 182301 bases at least Q20  
Insert size: 165000; agarose-fp  
Insert size: 182309; sum-of-contigs  
Quality coverage: 11.37x in Q20 bases; agarose-fp  
Quality coverage: 10.29x in Q20 bases; sum-of-contigs

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
consists of 5 contigs. Gaps between the contigs  
are represented as runs of N. The order of the pieces  
is believed to be correct as given, however the sizes  
of the gaps between them are based on estimates that have  
provided by the submittor.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
1  
24329: contig of 24329 bp in length  
24330: gap of unknown length  
24430: contig of 7193 bp in length  
31623: gap of unknown length  
31723: contig of 38989 bp in length  
70712: gap of unknown length  
70812: contig of 99994 bp in length  
170806: gap of unknown length  
170906: contig of 11804 bp in length.  
170906: 182709: contig of 11804 bp in length.

-----  
Location/Qualifiers  
1. .182709  
/organism="Sus scrofa"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9823"  
/clone="RP44-35K6"  
/clone\_lib="RP44"  
1. .38969  
/notes="Clone overlaps with GenBank Accession Number  
AC097631 clone RP44-62F17 (center project name cta)"  
1. .24329  
/notes="assembly\_fragment  
clone\_end:T7  
vector\_side:left"  
24330..24429  
/estimated\_length=unknown  
24430..31622  
/notes="assembly\_fragment"  
31623..31722  
/estimated\_length=unknown  
31723..70711  
/notes="assembly\_fragment"  
70712..70811  
/estimated\_length=unknown  
70812..170805  
/notes="assembly\_fragment"  
108812..182709  
/notes="clone overlaps with GenBank Accession Number  
AC097629 clone RP44-265A21 (center project name ctc)"  
170806..170905  
/estimated\_length=unknown  
170906..182709  
/notes="assembly\_fragment  
clone\_end:SP6  
vector\_side:right"

ORIGIN  
Query Match 91.6%; Score 17.4; DB 12; Length 182709;  
Best Local Similarity 94.7%; Pred. No. 3.5e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 CTTTACTTCATAGTCTTTG 19  
|||||  
Db 135412 CTTTACTTCATAGTCTTTG 135394

## RESULT 64

AC146837 183404 bp DNA linear HTG 30-OCT-2003  
 LOCUS Canis familiaris clone RP81-48L12, WORKING DRAFT SEQUENCE, 11  
 DEFINITION ordered pieces.

## ACCESSION

AC146837

## VERSION

HTG; HTGS PHASE2; HTGS DRAFT.

## KEYWORDS

Canis familiaris (dog)

## SOURCE

Canis familiaris

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;  
 Canis.

## REFERENCE

## AUTHORS

1 (bases 1 to 183404)  
 Antonellis, A., Ayele, K., Benjamin, B., Blakesley, R. W.,  
 Bouffard, G. G., Brinkley, C., Brooks, S., Chu, G., Coleman, B.,  
 Coleman, H., Engle, J., Granite, S., Guan, X., Gupta, J., Haghighi, P.,  
 Har, J., Hansen, N., Ho, S.-L., Hu, P., Hurle, B., Idol, J. R.,  
 Karlins, E., Kwong, P., Laric, P., Larson, S., Lee-Lin, S.-Q.,  
 Legaspi, R., Maduro, Q. L., Maduro, V. B., Margulies, E. H., Masiello, C.,  
 Maskeri, B., McDowell, J., Mullikin, J. C., Paquirigan, C., Pearson, R.,  
 Portnoy, M. E., Prasad, A., Reddix-Dugue, N., Schandler, K.,  
 Schueler, M. G., Shah, K., Sison, C., Stantripop, S., Thomas, J. W.,  
 Thomas, P. J., Tsipouri, V., Vogt, J. L., Wetherby, K. D., Young, A. and  
 Green, E. D.

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

\* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* provided by the submitter.

\* This sequence will be replaced

\* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.

\* 7223: contig of 7223 bp in length

\* 7323: gap of unknown length

\* 7324: contig of 23004 bp in length

\* 30328: gap of unknown length

\* 30428: contig of 5424 bp in length

\* 35851: gap of unknown length

\* 35852: contig of 13547 bp in length

\* 49498: gap of unknown length

\* 49499: contig of 36431 bp in length

\* 49539: gap of unknown length

\* 86030: contig of 7276 bp in length

\* 86130: gap of unknown length

\* 93406: gap of unknown length

\* 93505: contig of 42498 bp in length

\* 136003: gap of unknown length

\* 136004: contig of 20457 bp in length

\* 136104: gap of unknown length

\* 156561: gap of unknown length

\* 156661: contig of 2180 bp in length

\* 158941: gap of unknown length

\* 181753: contig of 22812 bp in length

\* 181852: gap of unknown length

\* 181853: contig of 1552 bp in length.

## FEATURES

## source

1. 183404

/organism="Canis familiaris"

/mol\_type="genomic DNA"

/db\_xref="taxon:9615"

/clone="RP81-48L12"

/clone\_lib="RP81"

1. 21343

/note="clone overlaps with GenBank Accession Number

AC146835 clone RP81-292E2 (center project name fyo)"

1. 7223

/note="assembly\_fragment

clone\_end:SP6

vector\_side:left"

7224. 7323

/estimated\_length=unknown

7324. 30327

/note="assembly\_fragment"

30328. 30427

/estimated\_length=unknown

30428. 35851

/note="assembly\_fragment"

35852. 35951

/estimated\_length=unknown

35952. 49498

/note="assembly\_fragment"

49499. 49598

/estimated\_length=unknown

49599. 86029

/note="assembly\_fragment"

86030. 86129

/estimated\_length=unknown

86130. 93405

/note="assembly\_fragment"

93406. 93505

/estimated\_length=unknown

93506. 136003

/note="assembly\_fragment"

136004. 136103

/estimated\_length=unknown

136104. 156560

/note="assembly\_fragment"

156561. 156660

/estimated\_length=unknown

156661. 158840

The sequence data in this record represents an 'enhanced'  
 version of a Phase 2 submission. Specifically, the indicated  
 order and orientation of each sequence contig has been  
 established using one or more of the following: read-pair  
 data from individual subclones, overlaps with neighboring  
 clones, alignment with available reference sequence (e.g.,  
 human), and/or confirmation by PCR testing. In addition,  
 the sequence assembly is based on at least 8X average  
 coverage in Q20 bases and has been reviewed to rule out  
 gross misassemblies, the low-quality ends of sequence  
 contigs have been trimmed away, and each base is associated  
 with a Phrap-derived quality score.

## Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 180485 bases at least Q40

Consensus quality: 181505 bases at least Q30

Consensus quality: 182100 bases at least Q20

Insert size: 176000; agarose-fp

Insert size: 182404; sum-of-contigs

Quality coverage: 13.58x in Q20 bases; agarose-fp

Quality coverage: 13.10x in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 11 contigs. Gaps between the contigs

```

/notes="assembly_fragment"
15841. .15840
/estimated_length=unknown
15841. .181752
/notes="assembly_fragment"
16657. .183404
/notes="clone overlaps with GenBank Accession Number
AC146771 clone RP81-121B12 (center project name fyl)"
181753. .181852
/estimated_length=unknown
181853. .183404
/notes="assembly_fragment
clone_end:T7
vector_side:right"

ORIGIN
Query Match          91.6%; Score 17.4; DB 12; Length 183404;
Best Local Similarity 94.7%; Pred. No. 3.5e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGCTCTTG 19
|||||
Db 165279 CATTACTTCATAGCTCTTG 165297

RESULT 65
AC011947/c 183812 bp DNA linear HTG 23-MAY-2001
LOCUS Homo sapiens clone RP11-17D12, WORKING DRAFT SEQUENCE, 6 unordered
DEFINITION Homo sapiens
ACCESSION AC011947
VERSION AC011947.4 GI:14190733
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 183812)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-17D12
Unpublished
2 (bases 1 to 183812)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Batra,N., Beckerly,R., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Castelle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessaye,S., Titrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (16-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 23, 2001 this sequence version replaced by:7387330.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L3625
Center clone name: 17.D.12
----- Summary Statistics

```

```

Sequencing vector: M13; M7815; 0% of reads
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 179876 bases at least Q40
Consensus quality: 182173 bases at least Q30
Consensus quality: 182976 bases at least Q20
Insert size: 183000; agarose-fp
Insert size: 183312; sum-of-contrigs
Quality coverage: 7.7 in Q20 bases; agarose-fp
Quality coverage: 7.7 in Q20 ba.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contrigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contrigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 81681: contrig of 81681 bp in length
* 81682 81781: gap of 100 bp
* 81782 83362: contrig of 1581 bp in length
* 83363 83462: gap of 100 bp
* 83463 84806: contrig of 1344 bp in length
* 84807 84906: gap of 100 bp
* 84907 85699: contrig of 793 bp in length
* 85700 85799: gap of 100 bp
* 85800 115070: contrig of 29271 bp in length
* 115071 115170: gap of 100 bp
* 115171 183812: contrig of 68642 bp in length.
FEATURES
source
1. .183812
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="RP11-17D12"
/clone_lib="RPC1-11 Human Male BAC"
misc_feature
1. .81681
/notes="assembly_fragment
clone_end:SP6
vector_side:left"
gap
81682. .81781
/estimated_length=100
misc_feature
81782. .83362
/notes="assembly_fragment"
gap
83363. .83462
/estimated_length=100
misc_feature
83463. .84806
/notes="assembly_fragment"
gap
84807. .84906
/estimated_length=100
misc_feature
84907. .85699
/notes="assembly_fragment"
gap
85700. .85799
/estimated_length=100
misc_feature
85800. .115070
/notes="assembly_fragment"
gap
115071. .115170
/estimated_length=100
misc_feature
115171. .183812
/notes="assembly_fragment
clone_end:T7
vector_side:right"
ORIGIN
Query Match          91.6%; Score 17.4; DB 12; Length 183812;
Best Local Similarity 94.7%; Pred. No. 3.5e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGCTCTTG 19
|||||
Db 100430 CTTTACTTCATAGCTCTTG 100412

```

```

RESULT 66
AC069476      184824 bp      DNA      linear      PRI 09-APR-2003
LOCUS      Homo sapiens chromosome 11 clone b430h11 map 11, complete sequence.
DEFINITION
AC069476
VERSION      AC069476.25 GI:29244703
KEYWORDS
SOURCE
ORGANISM      Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.
1 (bases 1 to 184824)
Deschamps,S., Richter,N. and Roe,B.A.
Homo sapiens Chromosome 11 BAC Clone b430h11
Unpublished
2 (bases 1 to 184824)
Deschamps,S., Richter,N. and Roe,B.A.
Direct Submission
Submitted (01-JUN-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
3 (bases 1 to 184824)
Deschamps,S., Richter,N. and Roe,B.A.
Direct Submission
Submitted (14-MAR-2003) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
4 (bases 1 to 184824)
Deschamps,S., Richter,N. and Roe,B.A.
Direct Submission
Submitted (19-MAR-2003) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
5 (bases 1 to 184824)
Deschamps,S., Richter,N. and Roe,B.A.
Direct Submission
Submitted (26-MAR-2003) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
6 (bases 1 to 184824)
Deschamps,S., Richter,N. and Roe,B.A.
Direct Submission
Submitted (09-APR-2003) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
On Mar 26, 2003 this sequence version replaced gi:28951189.
-----
Center: Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR
-----
FEATURES
source
Location/Qualifiers
1. 184824
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="11"
/map="11"
/clones="b430h11"
/clone_lib="RPCI Human Female PAC Library 6"
/note="This clone was originally contaminated with 10kb of
mollusc sequence"
ORIGIN
Query Match      91.6%; Score 17.4; DB 5; Length 184824;
Beat Local Similarity 94.7%; Pred. No. 3.5e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 CTTTACTTCATGCTTTG 19
Db 50197 CTTTACTTCATGCTTTG 50215

```

## RESULT 67

```

AC132501
LOCUS      Rattus norvegicus clone CH230-306G10, *** SEQUENCING IN PROGRESS
DEFINITION
AC132501
VERSION      AC132501.3 GI:25139146
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM      Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Rattus.
1 (bases 1 to 185079)
Muzny,D., Marie., Metzker,M., Lee., Abranzone,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Louissegh,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,S.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Naackelemehe,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Umanai,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wlezyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 185079)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (01-SEP-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

```

```

REFERENCE
AUTHORS      3 (bases 1 to 185079)
TITLE        Rat Genome Sequencing Consortium.
JOURNAL      Direct Submission
             Submitted (20-NOV-2002) Human Genome Sequencing Center, Department
             of Molecular and Human Genetics, Baylor College of Medicine, One
             Baylor Plaza, Houston, TX 77030, USA
COMMENT      On Nov 20, 2002 this sequence version replaced gi:22855833.
             The sequence in this assembly is a combination of BAC based reads
             and whole genome shotgun sequencing reads assembled using Atlas
             (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
             in the feature table below represents a scaffold in the Atlas
             assembly (a 'contig-scaffold'). Within each contig-scaffold,
             individual sequence contigs are ordered and oriented, and separated
             by sized gaps filled with Ns to the estimated size. The sequence
             may extend beyond the ends of the clone and there may be sequence
             contigs within a contig-scaffold that consist entirely of whole
             genome shotgun sequence reads. Both end sequences and whole genome
             shotgun sequence only contigs will be indicated in the feature
             table.
             ----- Genome Center
             Center: Baylor College of Medicine
             Center code: BCM
             Web site: http://www.hgsc.bcm.tmc.edu/
             Contact: hgsc-help@bcm.tmc.edu
             ----- Project Information
             Center project name: KBSN
             Center clone name: CH230-306G10
             ----- Summary Statistics
             Assembly program: Phrap; version 0.990329
             Consensus quality: 173513 bases at least Q40
             Consensus quality: 175951 bases at least Q30
             Consensus quality: 177426 bases at least Q20
             Estimated insert size: 176985; sum-of-contigs estimation
             Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
             -----
             * NOTE: Estimated insert size may differ from sequence length
             * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
             * NOTE: This is a 'working draft' sequence. It currently
             * consists of 3 contigs. The true order of the pieces
             * is not known and their order in this sequence record is
             * arbitrary. Gaps between the contigs are represented as
             * runs of N, but the exact sizes of the gaps are unknown.
             * This record will be updated with the finished sequence
             * as soon as it is available and the accession number will
             * be preserved.
             *
             * 1 29987: contig of 29987 bp in length
             * 29988 30087: gap of unknown length
             * 30088 183113: contig of 153026 bp in length
             * 183114 183213: gap of unknown length
             * 183214 185079: contig of 1866 bp in length.
             *
             * Location/Qualifiers
             1..185079
             /organism="Rattus norvegicus"
             /mol_type="genomic DNA"
             /db_xref="taxon:10116"
             /clone="CH230-306G10"
             29988..30087
             /estimated_length=unknown
             183114..183213
             /estimated_length=unknown

gap
gap

ORIGIN
Query Match          91.6%; Score 17.4; DB 12; Length 185079;
Best Local Similarity 94.7%; Pred. No. 3.5e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19
    ||||| ||||| |||||
Db 147719 CTTTAGTTCATAGTCTTTG 147737

RESULT 68
AC161349/c

```

```

LOCUS       AC161349               186281 bp    DNA    linear    HTG 12-MAY-2005
DEFINITION  Mus musculus chromosome 1 clone RP23-456J17 map 1, *** SEQUENCING
IN PROGRESS ***, 5 unordered pieces.
ACCESSION   AC161349
VERSION     AC161349.1  GI:63253447
KEYWORDS    HTG; HTGS PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 186281)
AUTHORS     Birren,B., Nusbaum,C. and Lander,E.
TITLE        Mus musculus chromosome 1, clone RP23-456J17
JOURNAL      Unpublished
REFERENCE   2 (bases 1 to 186281)
AUTHORS     Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,
            Bloom,T., Boguskiavkiy,L., Boukhgalter,B., Camarata,J., Chang,J.,
            Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,
            DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorrts,L.,
            Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D.,
            Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,
            Johnson,R., Jones,C., Kamat,A., Karatsa,A., Kells,C., Landers,T.,
            Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,
            MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,
            Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Stubbs,M., Talamas,J., Tenfaye,S., Theodore,J.,
            Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,
            Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
            Zimmer,A. and Zody,M.
            Direct Submission
            Submitted (12-MAY-2005) Broad Institute of MIT and Harvard, 320
            Charles Street, Cambridge, MA 02141, USA
            All repeats were identified using RepeatMasker:
            Smit,A.F.A. & Green,P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ----- Genome Center
            Center: Broad Institute of MIT and Harvard
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@broad.mit.edu
            ----- Project Information
            Center project name: L31746
            Center clone name: 456_J_17
            -----
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved
            *
            * 1 103107: contig of 103107 bp in length
            * 103108 103207: gap of unknown length
            * 103208 114179: contig of 10972 bp in length
            * 114180 114279: gap of unknown length
            * 114280 117610: contig of 3331 bp in length
            * 117611 117710: gap of unknown length
            * 117711 120716: contig of 3006 bp in length
            * 120717 120816: gap of unknown length
            * 120817 186281: contig of 65465 bp in length.
            *
            * Location/Qualifiers
            1..186281
            /organism="Mus musculus"
            /mol_type="genomic DNA"

```





CLONE LENGTH: This sequence represents the entire insert of this clone unless otherwise noted. If there are overlapping clones, the overlaps are noted in the beginning and end of the Features section.

## FEATURES

source

Location/Qualifiers

1. .189563  
/organism="Aotus nancymaae"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:37293"  
/clone="CH258-251O19"  
/clone\_lib="CH258"  
/note="BAC resource: <http://bacpac.chori.org/>"

## ORIGIN

Query Match 91.6%; Score 17.4; DB 5; Length 189563;  
Best Local Similarity 94.7%; Pred. No. 3.5e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 CTTTACTTCATAGCTTTG 19

Db 187640 CTTTACTTCATAGCTCTG 187622

## RESULT 71

AP001797

LOCUS

DEFINITION Homo sapiens chromosome 18 clone RP11-748120 map 18q12, WORKING  
DRAFT SEQUENCE, 27 unordered pieces.

ACCESSION AP001797

VERSION AP001797.2 GI:8117469

KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.

SOURCE Homo sapiens (human)

## ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Homnidae; Homo.

1 (bases 1 to 190550)

Hattori.M., Ishii.K., Toyoda.A., Taylor,T.D., Hong-Seog.P.,

Fujiyama.A., Yada.T., Totoki.Y., Watanabe.H. and Sakaki.Y.

Homo sapiens 190,550 genomic DNA of 18q12

Published Only in DataBase (2000)

2 (bases 1 to 190550)

Hattori.M., Ishii.K., Toyoda.A., Taylor,T.D., Hong-Seog.P.,

Fujiyama.A., Yada.T., Totoki.Y., Watanabe.H. and Sakaki.Y.

Direct Submission

Submitted (11-APR-2000) Masahira Hattori, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);

Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,

Japan (E-mail: [hattori@gsc.riken.go.jp](mailto:hattori@gsc.riken.go.jp),

URL: <http://hgp.gsc.riken.go.jp/>, Tel:81-42-778-9923,

Fax:81-42-778-9924)

On May 30, 2000 this sequence version replaced gi:7592912.

## COMMENT

----- Genome Center

Center: RIKEN Genomic Sciences Center(GSC)

Center code: RIKEN

Web site: <http://hgp.gsc.riken.go.jp/>

Contact: [hattori@gsc.riken.go.jp](mailto:hattori@gsc.riken.go.jp)

----- Project Information

Center project name: HumDraft18

Center clone name: RP11-748120

----- Summary Statistics

Sequencing vector: PCR products; 100% of reads

Chemistry: Dye-terminator ET-amersham; 100% of reads

Assembly program: Phrap; version 0.990129

Consensus quality: 178641 bases at least Q40

Consensus quality: 184743 bases at least Q30

Consensus quality: 186891 bases at least Q20

Insert size: 187950; sum-of-contigs

Quality coverage: 4.84x in Q20 bases; sum-of-contigs

-----

NOTE: This is a 'working draft' sequence. It currently consists of  
27 contigs. The true order of the pieces is not known and their

order in this sequence record is arbitrary. Gaps between the  
contigs are represented as runs N, but the exact sizes of the gaps  
are unknown. This record will be updated with the finished sequence  
as soon as it is available and the accession number will be  
preserved

1 16718 contig of 16718 bp in length  
16819 30449 contig of 13631 bp in length  
30550 45067 contig of 15058 bp in length  
45708 59752 contig of 14045 bp in length  
59853 75894 contig of 16042 bp in length  
75995 87034 contig of 11040 bp in length  
87135 97062 contig of 9928 bp in length  
97163 106435 contig of 9273 bp in length  
106536 114800 contig of 8265 bp in length  
114901 121600 contig of 6700 bp in length  
121701 129422 contig of 7722 bp in length  
129523 138872 contig of 9350 bp in length  
138973 145365 contig of 6393 bp in length  
145466 151207 contig of 5742 bp in length  
151308 155673 contig of 4366 bp in length  
155774 160475 contig of 5105 bp in length  
160576 165680 contig of 4702 bp in length  
165781 169630 contig of 3850 bp in length  
169731 174769 contig of 5039 bp in length  
174870 178072 contig of 3203 bp in length  
178173 180351 contig of 2179 bp in length  
180452 181295 contig of 844 bp in length  
181396 184552 contig of 3157 bp in length  
184653 186282 contig of 1630 bp in length  
186383 188196 contig of 1814 bp in length  
188397 189386 contig of 1090 bp in length  
189487 190550 contig of 1064 bp in length

Sequence updated (26-May-2000).

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 27 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 16718: contig of 16718 bp in length  
\* 16719 16818: gap of 100 bp  
\* 16819 30449: contig of 13631 bp in length  
\* 30450 30549: gap of 100 bp  
\* 30550 45607: contig of 15058 bp in length  
\* 45608 45707: gap of 100 bp  
\* 45708 59752: contig of 14045 bp in length  
\* 59753 59852: gap of 100 bp  
\* 59853 75894: contig of 16042 bp in length  
\* 75895 75994: gap of 100 bp  
\* 75995 87034: contig of 11040 bp in length  
\* 87035 87134: gap of 100 bp  
\* 87135 97062: contig of 9928 bp in length  
\* 97063 97162: gap of 100 bp  
\* 97163 106435: contig of 9273 bp in length  
\* 106436 106536: gap of 100 bp  
\* 106536 114800: contig of 8265 bp in length  
\* 114801 114900: gap of 100 bp  
\* 114901 121600: contig of 6700 bp in length  
\* 121601 121700: gap of 100 bp  
\* 121701 129422: contig of 7722 bp in length  
\* 129423 129522: gap of 100 bp  
\* 129523 138872: contig of 9350 bp in length  
\* 138873 138972: gap of 100 bp  
\* 138973 145365: contig of 6393 bp in length  
\* 145366 145465: gap of 100 bp  
\* 145466 151207: contig of 5742 bp in length  
\* 151208 151308: gap of 100 bp  
\* 151308 155673: contig of 4366 bp in length  
\* 155674 155773: gap of 100 bp  
\* 155774 160475: contig of 4702 bp in length  
\* 160476 160575: gap of 100 bp



```
AUTHORS
TITLE Direct Submission
JOURNAL Submitted (20-OCT-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 194492)
REFERENCE
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (26-MAY-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 194492)
REFERENCE
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (29-MAY-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
COMMENT
On May 29, 2002 this sequence version replaced gi:21206073.
INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the
entire insert of this clone. Overlapping regions of clones are only
sequenced and submitted once, so the sequence for the remainder of
the insert may be found in the record for the adjacent clones.
Overlapping clones are noted at the beginning and end of the
Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches
of a local database that includes entries from dbSTS, GDB, and
local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green,
unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
EST and cDNA sequences. Genes demonstrate at least two exons
flanked by consensus splice sites that maintained sequence
continuity across the splice junctions. Sequences that are not
identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum
standard of double strand coverage with a minimum of 2 clones and 2
reads with no ambiguities or 2 chemistries with a minimum of 2
clones and 3 reads with no ambiguities. If the sequence quality for
a region does not meet this standard, it will be indicated in the
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality
standards - estimated error rate less than 1 per 10,000 bases.
Reports of lowest quality individual bases and measures of base
quality are listed below. Description of the metrics can be found
at URL:
http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

FEATURES
source
1..194492
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/Chromosome="X"
/Clone="RP11-478H11"
88..242
/repeat_region
/rpt_family="AluJo/FRAM"
648..2719
/repeat_region
/rpt_family="L1M4A"
2720..3012
/repeat_region
/rpt_family="AluSx"
3013..4086
/repeat_region
/rpt_family="L1M4A"
4100..4396
/repeat_region
/rpt_family="AluJo"

complement(4786..5095)
/rpt_family="AluSg"
6104..6764
/rpt_family="L1ME3A"
6869..6980
/rpt_family="L2"
7028..7075
/rpt_family="(CA)n"
7076..7106
/rpt_family="(CA)n"
7178..7260
/rpt_family="(TA)n"
8004..8054
/rpt_family="(TTTC)n"
complement(8061..8194)
/rpt_family="FLAM_C"
8961..9258
/rpt_family="AluY"
11398..11450
/rpt_family="AT rich"
complement(11453..11520)
/rpt_family="MIR"
complement(11596..12790)
/rpt_family="L2"
12833..13015
/rpt_family="GA-rich"
13040..13661
/rpt_family="GA-rich"
13699..13881
/rpt_family="GA-rich"
14315..14612
/rpt_family="AluSx"
14616..14661
/rpt_family="(GAAA)n"
16525..16625
/rpt_family="MIR"
16850..17414
/rpt_family="LTR55"
17828..17950
/rpt_family="MIR"
18539..18575
/rpt_family="MIR"
18576..18881
/rpt_family="MER33"
18882..18922
/rpt_family="MIR"
complement(19026..19064)
/rpt_family="L2"
complement(19064..19207)
/rpt_family="MIR"
19371..19397
/rpt_family="(TC)n"
complement(19726..20140)
/rpt_family="L1ME"
21404..22846
/rpt_family="Tigger1"
22874..22947
/rpt_family="(TTAA)n"
22968..23906
/rpt_family="Tigger1"
24713..24922
/rpt_family="L1PA6"
complement(25204..25514)
/rpt_family="AluY"
25758..25949
/rpt_family="MER30"
26165..26240
/rpt_family="L2"
complement(26692..27003)
/rpt_family="AluY"
28270..28398
/rpt_family="L2"
complement(29047..29258)
```

```

Query Match          91.6%; Score 17.4; DB 5; Length 194492;
Best Local Similarity 94.7%; Pred. No. 3.6e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19
    ||||| ||||| |||||
Db 29223 CTTTACTCCATAGTCTTTG 29205

RESULT 73
LOCUS AC122492/c
DEFINITION Mus musculus BAC clone RP24-390N7 from chromosome 5, complete
sequence.
ACCESSION AC122492
VERSION AC122492.3 GI:28209796
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.
1 (bases 1 to 196275)
Harkins, R., Cotton, M., Haglund, K. and Bielicki, L.
The sequence of Mus musculus BAC clone RP24-390N7
Unpublished (2001)
2 (bases 1 to 196275)
Wilson, R.
Sequencing of Mus musculus
Unpublished (2001)
3 (bases 1 to 196275)
McPherson, J.D. and Waterston, R.H.
Direct Submission
Submitted (23-MAY-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
4 (bases 1 to 196275)
McPherson, J.D. and Waterston, R.H.
Direct Submission
Submitted (22-JUN-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
5 (bases 1 to 196275)
McPherson, J.D. and Waterston, R.H.
Direct Submission
Submitted (05-FEB-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
6 (bases 1 to 196275)
Wilson, R.
Direct Submission
Submitted (13-NOV-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Feb 5, 2003 this sequence version replaced gi:21539185.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@watson.wustl.edu
----- Summary Statistics
----- Center project name: M_BB0390N07
-----

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

#### SOURCE INFORMATION:

The RPCI-24 BAC Library has been constructed by Pieter de Jong and coworkers (<http://www.chori.org>) from male C57BL/6J mouse spleen and/or brain genomic DNA. The clone and detailed information can be obtained from Pieter de Jong and coworkers at <http://www.chori.org>

#### NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone. This clone is overlapped by AC115118.

FEATURES	source
	Location/Qualifiers
	1..196275
	/organism="Mus musculus"
	/mol_type="genomic DNA"
	/db_xref="taxon:10090"
	/chromosome="5"
	/map="5"
	/clone="RP24-390N7"
	/clone_lib="RPCI-24"
repeat_region	38..369
	/rpt_family="MaLR"
repeat_region	370..964
	/rpt_family="L1"
repeat_region	963..1419
	/rpt_family="L1"
repeat_region	2316..2473
	/rpt_family="L1"
repeat_region	4038..4089
	/rpt_family="B4"
repeat_region	4104..4214
	/rpt_family="RMR1B"
repeat_region	4394..4523
	/rpt_family="Alu"
repeat_region	4798..5009
	/rpt_family="B2"
repeat_region	5154..5333
	/rpt_family="B2"
repeat_region	7461..8097
	/rpt_family="L1"
repeat_region	8534..9541
	/rpt_family="L1"
repeat_region	9578..9837
	/rpt_family="L1"
repeat_region	9860..10091
	/rpt_family="L1"
repeat_region	10092..10518
	/rpt_family="ERVK"
repeat_region	10519..10703
	/rpt_family="L1"
repeat_region	11526..12174
	/rpt_family="L1"
repeat_region	12559..13158
	/rpt_family="L1"
repeat_region	13158..13418
	/rpt_family="L1"
repeat_region	13407..13834
	/rpt_family="L1"
repeat_region	13831..14876
	/rpt_family="L1"
repeat_region	14877..15272
	/rpt_family="MaLR"
repeat_region	15273..15311
	/rpt_family="L1"
repeat_region	15310..16540
	/rpt_family="L1"
repeat_region	16944..17274
	/rpt_family="L1"

```
repeat_region 17275. .17459
/rpt_family="B2"
tRNA 17280. .17353
/product="tRNA-Ser"
/notes="Likely pseudogene (HMM SC=42.36 / Sec struct
SC=22.25)"
repeat_region 17460. .17685
/rpt_family="L1"
repeat_region 17686. .17784
/rpt_family="L1"
repeat_region 17799. .20144
/rpt_family="MaLR"
repeat_region 20913. .21432
/rpt_family="ERVK"
repeat_region 22460. .22804
/rpt_family="MaLR"
repeat_region 24845. .25333
/rpt_family="L1"
repeat_region 25777. .25827
/rpt_family="ERV1"
repeat_region 26215. .26709
/rpt_family="L1"
repeat_region 26745. .27207
/rpt_family="L1"
repeat_region 29209. .29949
/rpt_family="L1"
repeat_region 31048. .31222
/rpt_family="B2"
repeat_region 32006. .32407
/rpt_family="L1"
repeat_region 32442. .32788
/rpt_family="L1"
repeat_region 33915. .34069
/rpt_family="B4"
repeat_region 34573. .35037
/rpt_family="L1"
repeat_region 35872. .35927
/rpt_family="ERV1"
repeat_region 36012. .36163
/rpt_family="L1"
repeat_region 36323. .36553
/rpt_family="L1"
repeat_region 37016. .37126
/rpt_family="Alu"
repeat_region 37425. .37616
/rpt_family="MER1_type"
repeat_region 39478. .39609
/rpt_family="ID"
repeat_region 40439. .40836
/rpt_family="L1"
repeat_region 40994. .41098
/rpt_family="L1"
repeat_region 41197. .41448
/rpt_family="MaLR"
repeat_region 43283. .43551
/rpt_family="L1"
repeat_region 43790. .43979
/rpt_family="B4"
repeat_region 44009. .44154
/rpt_family="Alu"
repeat_region 44156. .44271
/rpt_family="B4"
repeat_region 44664. .45063
/rpt_family="ERVK"
repeat_region 48337. .49516
/rpt_family="ERVK"
Query Match 91.6%; Score 17.4; DB 6; Length 196275;
Best Local Similarity 94.7%; Pred. No. 3.6e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 CTTTACTTCATAGTCTTTG 19
||||| ||||||| |||||||
```

```
Db 5924 CTTTCTTCATAGTCTTTG 5906
RESULT 74
AC097629/c
LOCUS AC097629.2 196700 bp DNA linear HTG 13-MAR-2002
DEFINITION Sus scrofa clone RP44-265A21, WORKING DRAFT SEQUENCE, 5 ordered
pieces.
ACCESSION AC097629
VERSION GI:19387618
KEYWORDS HTG: HTGS PHASE2; HTGS_DRAFT.
SOURCE Sus scrofa (pig)
ORGANISM Sus.
REFERENCE 1 (bases 1 to 196700)
AUTHORS Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C.,
Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J.,
Haghghi,P., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E., Laric,P.,
Lee-Lin,S.-O., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
Margulies,E.H., Masiello,C., Maskeri,B., Mastrian,S.D.,
McCloskey,J.C., McDowell,J., Paguirigan,C., Pearson,R.,
Portnoy,M.E., Prasad,A., Schueler,M.G., Stantriop,S., Thomas,J.W.,
Thomas,P.J., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A.,
Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 196700)
AUTHORS Green,E.D.
TITLE Direct Submission
JOURNAL Submitted (20-OCT-2001) NIH Intramural Sequencing Center, 8717
Grovmont Circle, Gaithersburg, MD 20877, USA
REFERENCE 3 (bases 1 to 196700)
AUTHORS Green,E.D.
TITLE Direct Submission
JOURNAL Submitted (13-MAR-2002) NIH Intramural Sequencing Center, 8717
Grovmont Circle, Gaithersburg, MD 20877, USA
On Mar 13, 2002 this sequence version replaced gi:16303428.
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc.mouse@nhgri.nih.gov
----- Project Information
Center project name: ctc
Center clone name: 265A21
The sequence data in this record represents an 'enhanced'
version of a Phase 2 submission. Specifically, the indicated
order and orientation of each sequence contig has been
established using one or more of the following: read-pair
data from individual subclones, overlaps with neighboring
clones, alignment with available reference sequence (e.g.,
human), and/or confirmation by PCR testing. In addition,
the sequence assembly is based on at least 8x average
coverage in Q20 bases and has been reviewed to rule out
gross misassemblies, the low-quality ends of sequence
contigs have been trimmed away, and each base is associated
with a Phrap-derived quality score.
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 195502 bases at least Q40
Consensus quality: 196012 bases at least Q30
Consensus quality: 196230 bases at least Q20
Insert size: 174000; agarose-fp
Insert size: 196300; sum-of-contigs
Quality coverage: 10.63x in Q20 bases; agarose-fp
Quality coverage: 9.42x in Q20 bases; sum-of-contigs
-----
```

\* NOTE: This is a 'working draft' sequence. It currently consists of 5 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

\* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.  
 \* 1 61967: contig of 61967 bp in length  
 \* 61968 62067: gap of unknown length  
 \* 62068 128700: contig of 66633 bp in length  
 \* 128701 128800: gap of unknown length  
 \* 128801 144849: contig of 16049 bp in length  
 \* 144850 144949: gap of unknown length  
 \* 144950 169039: contig of 24090 bp in length  
 \* 169040 169139: gap of unknown length  
 \* 169140 196700: contig of 27561 bp in length.

## FEATURES

Location/Qualifiers  
 1. 196700  
 /organism="Sus scrofa"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9823"  
 /clone="RP44-265A21"  
 /clone\_lib="RP44"

1. 73532  
 /note="clone overlaps with GenBank Accession Number AC097585 clone RP44-35K6 (center project name ctb)"

## misc\_feature

1. 61967  
 /note="assembly\_fragment  
 clone end:T7  
 vector\_side:left"

## gap

61968..62067  
 /estimated\_length=unknown

## misc\_feature

62068..128700  
 /note="assembly\_fragment"

## misc\_feature

123997..196700  
 /note="clone overlaps with GenBank Accession Number AC098815 clone RP44-213D24 (center project name ctd)"

## gap

128701..128800  
 /estimated\_length=unknown

## misc\_feature

128801..144849  
 /note="assembly\_fragment"

## gap

144850..144949  
 /estimated\_length=unknown

## misc\_feature

144950..169039  
 /note="assembly\_fragment"

## gap

169040..169139  
 /estimated\_length=unknown

## misc\_feature

169140..196700  
 /note="assembly\_fragment  
 clone end:SP6  
 vector\_side:right"

## ORIGIN

Query Match 91.6%; Score 17.4; DB 12; Length 196700;

Best Local Similarity 94.7%; Pred. No. 3.6e+02; Gaps 0; Indels 0; Mismatches 1; Gaps 0;

1 CTTTACTTCATAGCTTTG 19

26597 CTTTACTTCATCTTTG 26579

## RESULT 75

AC102703 AC102703 197439 bp DNA linear ROD 29-APR-2005

LOCUS Mus musculus chromosome 1, clone RP24-442E5, complete sequence.

DEFINITION Mus musculus chromosome 1, clone RP24-442E5, complete sequence.

ACCESSION AC102703

VERSION AC102703.16 GI:62955003

KEYWORDS HTG.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.

1. (bases 1 to 197439)

Birren,B., Nusbaum,C. and Lander,E.

Mus musculus chromosome 1, clone RP24-442E5

Unpublished

2. (bases 1 to 197439)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B., Brown,A., Camarata,J., Campiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G., MacLean,C., MacDonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Straus,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

3. (bases 1 to 197439)

Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R., MacLean,C., MacDonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupack,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (18-MAR-2005) Broad Institute of MIT and Harvard, 320 Charles Street, Cambridge, MA 02141, USA

4. (bases 1 to 197439)

Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R., MacLean,C., MacDonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,

Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

# TITLE

## JOURNAL

Submitted (29-APR-2005) Broad Institute of MIT and Harvard, 320

Charles Street, Cambridge, MA 02141, USA

On Apr 29, 2005 this sequence version replaced gi:61651929.

All repeats were identified using RepeatMasker:

Smith, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Broad Institute of MIT and Harvard

Center code: W18R

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@broad.mit.edu](mailto:sequence_submissions@broad.mit.edu)

----- Project Information

Center project name: L19314

Center clone name: 442\_E\_5

## FEATURES

source

Location/Qualifiers

1. .197439  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="1"  
/map="1"  
/clone="RP24-442E5"  
/clone\_lib="RPCI-24 Male Mouse BAC"  
complement(1. .1725)  
/rpt\_family="L1"  
1. 634  
/note="PCR product sequence only"

repeat\_region

unsure

1. .112  
/note="single clone coverage"

unsure

complement(582. .634)

unsure

/note="single clone coverage"

unsure

complement(1422. .1527)

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region  
/rpt\_family="AT\_rich"  
18474. .18533  
/rpt\_family="(CATA)n"  
complement(18536. .18626)  
/rpt\_family="L1MC4"  
19529. .19554  
/rpt\_family="(TG)n"  
20605. .20648  
/rpt\_family="(TTTTG)n"  
complement(20661. .20802)  
/rpt\_family="B1\_MM"  
21732. .21765  
/rpt\_family="(CA)n"  
21770. .21836  
/rpt\_family="(TATATG)n"  
22515. .22654  
/rpt\_family="B1P"  
22657. .25314  
/rpt\_family="L1\_MM"  
complement(26335. .26508)  
/rpt\_family="B3"  
complement(26723. .32928)  
/rpt\_family="L1\_MM"  
27213. .27245  
/note="single clone coverage"  
27216. .27221  
/note="<30 Qual SNGL region"  
34104. .34158  
/rpt\_family="(CA)n"  
complement(34159. .34472)  
/rpt\_family="MTD"  
34507. .34840  
/rpt\_family="IAPLTR1\_MM"  
34907. .37169  
/rpt\_family="IAPLTR1\_MM-int"  
37169. .37224  
/rpt\_family="IAPLTR1\_MM-int"  
37223. .37579  
/rpt\_family="IAPLTR1\_MM-int"  
37571. .38345  
/rpt\_family="IAPLTR1\_MM-int"  
38348. .38681  
/rpt\_family="IAPLTR1\_MM"  
complement(38752. .39726)  
/rpt\_family="Lx8"  
complement(40083. .40510)  
/rpt\_family="L1MA5"

Query Match 91.6%; Score 17.4; DB 6; Length 197439;  
Best Local Similarity 94.7%; Pred. No. 3.6e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGCTCTTG 19

Db 146317 CTTTCTTCATAGCTCTTG 146335

## RESULT 76

AC115118  
LOCUS AC115118 198448 bp DNA linear ROD 05-NOV-2003  
DEFINITION Mus musculus BAC clone RP23-84N24 from 5, complete sequence.  
AC115118  
ACCESSION  
VERSION AC115118.3 GI:22138690  
KEYWORDS HTG.

SOURCE

ORGANISM

Mus musculus (house mouse)

Mus musculus

Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 198448)

ABSTRACT

Abbott, S., Haakenson, W. and Doebber, A.

TITLE The sequence of Mus musculus BAC clone RP23-84N24

JOURNAL Unpublished (2001)







## ORIGIN

Query Match 91.6%; Score 17.4; DB 12; Length 210493;  
 Best Local Similarity 94.7%; Pred. No. 3.6e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Oy 1 CTTTACTTCATAGTCTTTG 19  
 |||||  
 Db 24120 CTTTACTTCATAGTCTCTG 24102

## RESULT 78

AC113245/c  
 LOCUS AC113245 213428 bp DNA linear ROD 20-DEC-2005  
 DEFINITION Mus musculus chromosome 7, clone RP23-175D5, complete sequence.  
 ACCESSION AC113245  
 VERSION AC113245.12 GI:83745255  
 KEYWORDS HTG.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muroidae; Muridae; Murinae; Mus.  
 1 (bases 1 to 213428)  
 Birren, B., Nusbaum, C. and Lander, E.  
 Mus musculus chromosome 7, clone RP23-175D5  
 Unpublished

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,  
 Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B.,  
 Brown, A., Camarata, J., Campoliano, A., Chang, J., Chazaro, B.,  
 Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,  
 Cooke, P., Dearellano, K., Dewar, K., Diaz, J., Dodge, S., Faro, S.,  
 Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,  
 Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
 Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
 Kamat, A., Karatas, A., Kells, C., Larocque, K., Lamazares, R.,  
 Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C.,  
 Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M.,  
 McEwan, P., McKernan, K., Meldrim, J., Meneus, L., Mihova, T.,  
 Mieng, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,  
 Norman, C. H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J.,  
 Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,  
 Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,  
 Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S.,  
 Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
 Straus, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
 Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,  
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,  
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (27-FEB-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 213428)

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,  
 Anderson, M., Anderson, S., Arachchi, H. M., Barna, N., Bastien, V.,  
 Bloom, T., Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J.,  
 Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B.,  
 Dearellano, K., Diaz, J., Dodge, S., Dooley, K., Dorris, L.,  
 Erickson, J., Faro, S., Ferreira, P., FitzGerald, M., Gage, D.,  
 Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N.,  
 Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I.,  
 Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T.,  
 Levine, R., Lindblad-Toh, K., Liu, G., Liu, X., Lui, A., Mabbitt, R.,  
 MacLean, C., Macdonald, P., Major, J., Manning, J., Matthews, C.,  
 McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mieng, V.,  
 Murphy, T., Naylor, J., Nguyen, C., Nguyen, T., Nicol, R., Norbu, C.,  
 O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K.,  
 Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C.,  
 Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schuback, R.,  
 Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,  
 Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J.,  
 Topham, K., Travers, M., Vassiliev, H., Venkataraman, V. S., Viel, R.,  
 Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L.,  
 Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (19-FEB-2005) Broad Institute of MIT and Harvard, 320  
 Charles Street, Cambridge, MA 02141, USA  
 4 (bases 1 to 213428)

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## REFERENCE

## AUTHORS

## REFERENCE

## AUTHORS

## REFERENCE

## AUTHORS

## REFERENCE

## AUTHORS

## REFERENCE

## AUTHORS

## REFERENCE

## AUTHORS

## REFERENCE

## AUTHORS

## REFERENCE

## AUTHORS

## REFERENCE

## AUTHORS

## REFERENCE

## AUTHORS

## REFERENCE

## AUTHORS

## REFERENCE

## AUTHORS

## REFERENCE

## AUTHORS

## REFERENCE

## AUTHORS

## REFERENCE

## AUTHORS

## REFERENCE

## AUTHORS

## REFERENCE

## AUTHORS

## REFERENCE

## AUTHORS

## REFERENCE

## AUTHORS

## REFERENCE

## AUTHORS

## REFERENCE

## AUTHORS

## REFERENCE

## AUTHORS

## REFERENCE

## AUTHORS

## REFERENCE

## AUTHORS

## REFERENCE

## AUTHORS

## REFERENCE

## AUTHORS

## REFERENCE

## AUTHORS

## REFERENCE

## AUTHORS

## REFERENCE

## AUTHORS

## REFERENCE

## AUTHORS

## REFERENCE

## AUTHORS

## REFERENCE

## AUTHORS

## REFERENCE

## AUTHORS

## REFERENCE

## AUTHORS

## REFERENCE

```
repeat_region /chromosome="7"
repeat_region /map="7"
repeat_region /clone="RP23-175D5"
repeat_region /clone_lib="RPCI-23 Female Mouse BAC"
repeat_region 275..451
repeat_region /rpt_family="B3"
repeat_region 452..487
repeat_region /rpt_family="(TA)n"
repeat_region 488..532
repeat_region /rpt_family="(TG)n"
repeat_region 569..593
repeat_region /rpt_family="AT_rich"
repeat_region 594..696
repeat_region /rpt_family="L1M3"
repeat_region /rpt_family="B1_Mus1"
repeat_region 700..791
repeat_region 811..971
repeat_region /rpt_family="GA-rich"
repeat_region 967..1028
repeat_region /rpt_family="(GAAAA)n"
repeat_region 1155..1459
repeat_region /rpt_family="L1M3"
repeat_region complement(1500..1617)
repeat_region /rpt_family="MLT1A1"
repeat_region 1725..1818
repeat_region /rpt_family="B1F1"
repeat_region 1825..1938
repeat_region /rpt_family="GA-rich"
repeat_region 2221..2371
repeat_region /rpt_family="L1M3"
repeat_region 3244..3314
repeat_region /rpt_family="(TG)n"
repeat_region complement(3361..3424)
repeat_region /rpt_family="RSINE1"
repeat_region 3686..4061
repeat_region /rpt_family="Lx3B"
repeat_region complement(5077..5519)
repeat_region /rpt_family="RMER4A"
repeat_region 5537..5690
repeat_region /rpt_family="L1MD"
repeat_region 5720..6079
repeat_region /rpt_family="L1MC"
repeat_region 6152..6215
repeat_region /rpt_family="(TG)n"
repeat_region complement(6237..6635)
repeat_region /rpt_family="L1MA7"
repeat_region 6735..6776
repeat_region /rpt_family="AT_rich"
repeat_region complement(6782..7325)
repeat_region /rpt_family="L1M3"
repeat_region 7401..8107
repeat_region /rpt_family="L1M2"
repeat_region 8150..8546
repeat_region /rpt_family="L1M2"
repeat_region complement(8549..8715)
repeat_region /rpt_family="Lx9"
repeat_region 8801..9198
repeat_region /rpt_family="RLTR9A2"
repeat_region complement(9199..9339)
repeat_region /rpt_family="Lx9"
repeat_region 9345..9689
repeat_region /rpt_family="Lx9"
repeat_region 9698..9809
repeat_region /rpt_family="MTEa"
repeat_region 9855..10200
repeat_region /rpt_family="Lx9"
repeat_region complement(10249..10395)
```

Query Match 91.6%; Score 17.4; DB 6; Length 213428;  
Best Local Similarity 94.7%; Pred. No. 3.6e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19

```
Db 165304 CTTTAAATTCATAGTCTTTG 165286
||||| ||||| ||||| ||||| |||||
AC146883 216802 bp DNA linear PRI 21-JUN-2005
Callithrix jacchus clone CH259-225J7, complete sequence.
AC146883
AC146883.3 GI:68051757
HTG
Callithrix jacchus (white-tufted-ear marmoset)
Callithrix jacchus
Mammalia; Eutheria; Euarchontoglires; Primates; Platyrrhini;
Callitrichidae; Callithrix.
1 (bases 1 to 216802)
Antonellis,A., Ayele,K., Bass,D., Benjamin,B., Bera,J.,
Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S., Chu,G.,
Coleman,H., Engle,J., Fuxsenko,T., Gestole,M., Greene,A., Guan,X.,
Gupta,J., Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Hu,P.,
Hunter,G., Hurlie,B., Idol,J.R., Kwong,P., Laric,P., Larson,S.,
Lee-Lin,S.-O., Legaspi,R., Madden,M., Maduro,Q.L., Maduro,V.B.,
Margulies,E.H., Masiello,C., Maskeri,B., McDowell,J., Mojidi,H.A.,
Mullikin,J.C., Oestreicher,J.S., Park,M., Portnoy,M.E., Prasad,A.,
Puri,O., Reddix-Dugue,N., Sance,A., Schandler,K., Schueler,M.G.,
Sison,C., Stantripop,S., Stephen,E., Taye,A., Thomas,J.W.,
Thomas,P.J., Tsipouri,V., Ung,L., Vogt,J.L., Wetherby,K.D.,
Withers,T.R., Young,A. and Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 216802)
Green,E.D.
Direct Submission
Submitted (16-OCT-2003) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA
3 (bases 1 to 216802)
Green,E.D.
Direct Submission
Submitted (31-OCT-2003) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA
4 (bases 1 to 216802)
Green,E.D.
Direct Submission
Submitted (21-JUN-2005) NIH Intramural Sequencing Center, 5625
Fishers Lane, Rockville, MD 20852, USA
On Jun 21, 2005 this sequence version replaced gi:38093728.
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc.zoo@nhgri.nih.gov
----- Project Information
Center project name: ftb
Center Clone name: 225J07
```

This sequence was finished as follows unless otherwise noted:  
all regions were double-stranded, sequenced with an  
alternate chemistry, or covered by high quality data  
(i.e., phred quality >= 30); an attempt was made to resolve  
all sequencing problems, such as compressions and repeats;  
all regions were covered by at least one plasmid subclone  
or more than one M13 subclone; and the assembly was confirmed  
by restriction digest.

CLONE LENGTH: This sequence represents the entire insert of  
this clone unless otherwise noted. If there are overlapping  
clones, the overlaps are noted in the beginning and end of  
the features section.

FEATURES  
source  
1..216802  
/organism="Callithrix jacchus"  
/mol\_type="genomic DNA"

```

/db_xref="taxon:9483"
/clone="CH259-225J7"
/clone_lib="CH259"
/note="BAC resource: http://bacpac.chori.org/"
1938. .31985
/note="PCR product sequence only"

ORIGIN
-----
      91.6%; Score 17.4; DB 5; Length 216802;
      Best Local Similarity 94.7%; Pred. No. 3.6e+02;
      Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 CTTTACTTCATAGTCTTTG 19
    |||||
Db 143520 CTTTACTTCATAGTCTG 143502

RESULT 80
AC147845/c
LOCUS AC147845 217669 bp DNA linear HTG 24-FEB-2004
DEFINITION Saimiri boliviensis boliviensis clone CH254-335K11, WORKING DRAFT
SEQUENCE, 7 ordered pieces.
ACCESSION AC147845
VERSION AC147845.2 GI:42538994
KEYWORDS HTG; HTGS PHASE2; HTGS DRAFT.
SOURCE Saimiri boliviensis boliviensis (Bolivian squirrel monkey)
ORGANISM Saimiri boliviensis boliviensis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Platyrrhini;
Cebidae; Cebinae; Saimiri.
1 (bases 1 to 217669)
Antionellis,A., Ayele,K., Benjamin,B., Blakesley,R.W.,
Bouffard,G.G., Brinkley,C., Brooks,S., Chu,G., Coleman,B.,
Coleman,H., Daki,N., Engle,J., Granite,S., Guan,X., Gupta,J.,
Haghghi,P., Han,J., Hansen,N., Ho,S.-L., Hu,P., Hurle,B.,
Idol,J.R., Jones,C., Karlins,B., Kim,H., Kwong,P., Laric,P.,
Larson,S., Lee-lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
Margulies,E.H., Masiello,C., Maskeri,B., McDowell,J.,
Mullikin,J.C., Paguirigan,C., Portnoy,M.E., Prasad,A., Puri,O.,
Reddy-Dugue,N., Schander,K., Schueler,M.G., Shah,K., Sison,C.,
Scantripp,S., Thomas,J.W., Thomas,P.J., Tsipouri,V., Vogt,J.L.,
Wetherby,K.D., Young,A. and Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 217669)
Green,E.D.
Direct Submission
Submitted (15-JAN-2004) NIH Intramural Sequencing Center, 8717
Grovermont Circle, Gaithersburg, MD 20877, USA
3 (bases 1 to 217669)
Green,E.D.
Direct Submission
Submitted (12-FEB-2004) NIH Intramural Sequencing Center, 8717
Grovermont Circle, Gaithersburg, MD 20877, USA
On Feb 12, 2004 this sequence version replaced gi:40882645.
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc.zoo@nih.gov
----- Project Information
Center project name: gud
Center clone name: 335K11

The sequence data in this record represents an 'enhanced'
version of a Phase 2 submission. Specifically, the indicated
order and orientation of each sequence contig has been
established using one or more of the following: read-pair
data from individual subclones, overlaps with neighboring
clones, alignment with available reference sequence (e.g.,
human), and/or confirmation by PCR testing. In addition,
the sequence assembly is based on at least 8X average
coverage in Q20 bases and has been reviewed to rule out

```

gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 215445 bases at least Q40  
 Consensus quality: 216320 bases at least Q30  
 Consensus quality: 216749 bases at least Q20  
 Insert size: 203000; agarose-fp  
 Quality coverage: 9.44x in Q20 bases; agarose-fp  
 Quality coverage: 8.83x in Q20 bases; sum-of-contigs

-----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 7 contigs. Gaps between the contigs  
 \* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* provided by the submitter.

\* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.

1 51328: contig of 51328 bp in length  
 51329 51428: gap of unknown length  
 51429 134408: contig of 82980 bp in length  
 134409 134508: gap of unknown length  
 134509 178461: contig of 43953 bp in length  
 178462 178562: gap of unknown length  
 178563 184771: contig of 6210 bp in length  
 184772 184871: gap of unknown length  
 184872 190259: contig of 5388 bp in length  
 190260 190359: gap of unknown length  
 190360 216319: contig of 25960 bp in length  
 216320 216419: gap of unknown length  
 216420 217669: contig of 1250 bp in length.

#### FEATURES

##### source

1..217669  
 /organism="Saimiri boliviensis boliviensis"  
 /mol\_type="genomic DNA"  
 /sub\_species="boliviensis"  
 /db\_xref="taxon:39432"  
 /clone="CH254-335K11"  
 /clone\_lib="CH254"  
 /note="BAC resource: http://bacpac.chori.org/"

1..51328

##### misc\_feature

/note="assembly\_fragment"  
 clone end:T7  
 vector\_side:left  
 51329..51428

##### gap

/estimated\_length=unknown  
 51429..134408

##### misc\_feature

/note="assembly\_fragment"  
 134409..134508

##### gap

/estimated\_length=unknown  
 134509..178461

##### misc\_feature

/note="assembly\_fragment"  
 178462..178561

##### gap

/estimated\_length=unknown  
 178562..184771

##### misc\_feature

/note="assembly\_fragment"  
 184772..184871

##### gap

/estimated\_length=unknown  
 184872..190259

##### misc\_feature

/note="assembly\_fragment"  
 190260..190359

##### gap

/estimated\_length=unknown  
 190360..216319

##### misc\_feature

/note="assembly\_fragment"  
 20723..217669

/note="clone overlaps with GenBank Accession Number  
 AC147932 clone CH254-331B16 (center project name guc)"

```
gap          216320..216419
             /estimated_length=unknown
misc_feature 216420..217669
             /note="assembly_fragment
             clone_end:SP6
             vector_side:right"
ORIGIN
Query Match      91.6%; Score 17.4; DB 12; Length 217669;
Best Local Similarity 94.7%; Pred. No. 3.6e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy      1 CTTTACTTCATAGTCCTTTG 19
         |||||
Db      135502 CTTTACTTCATAGTCCTG 135484

RESULT 81
AC092434/c
LOCUS       AC092434      218872 bp DNA linear HTG 27-JUL-2001
DEFINITION Homo sapiens chromosome 4 clone RP11-210L21, *** SEQUENCING IN
            PROGRESS ***, 40 unordered pieces.
ACCESSION   AC092434
VERSION     AC092434.1 GI:14595956
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homiidae; Homo.
REFERENCE   1 (bases 1 to 218872)
            Waterston,R.H.
            The sequence of Homo sapiens clone
            Unpublished
REFERENCE   2 (bases 1 to 218872)
            Waterston,R.H.
            Direct Submission
            Submitted (04-JUL-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H NH0210L21
----- Summary Statistics -----
Sequencing vector: M13; 14%
Sequencing vector: plasmid; 86%
Chemistry: Dye-primer ET; 14% of reads
Chemistry: Dye-terminator Big Dye; 86% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 188006 bases at least Q40
Consensus quality: 196601 bases at least Q30
Consensus quality: 202813 bases at least Q20
Insert size: 181000; agarose-fp
Insert size: 214972; sum-of-contigs
Quality coverage: 7.25 in Q20 bases; agarose-fp
Quality coverage: 5.64 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 40 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1      2669: contig of 2669 bp in length
*      2670: gap of unknown length
*      2770 4554: contig of 1785 bp in length
*      4555 4654: gap of unknown length
```

```
*      4655 6908: contig of 2254 bp in length
*      6909 7008: gap of unknown length
*      7009 10387: contig of 3379 bp in length
*      10388 10487: gap of unknown length
*      10488 12825: contig of 2338 bp in length
*      12826 12925: gap of unknown length
*      12926 15663: contig of 2738 bp in length
*      15664 15763: gap of unknown length
*      15764 18992: contig of 3229 bp in length
*      18993 21894: contig of 2792 bp in length
*      21895 25307: contig of 3323 bp in length
*      25308 38126: contig of 12719 bp in length
*      38127 38226: gap of unknown length
*      38227 67916: contig of 29590 bp in length
*      67917 175443: contig of 107527 bp in length
*      175444 175544: gap of unknown length
*      175544 176879: contig of 1335 bp in length
*      176879 176979: gap of unknown length
*      176979 178261: contig of 1282 bp in length
*      178261 178361: contig of 1256 bp in length
*      178361 179617: gap of unknown length
*      179617 181151: contig of 1435 bp in length
*      181151 181252: gap of unknown length
*      181252 182972: contig of 1720 bp in length
*      182972 18476: contig of unknown length
*      18476 184577: contig of 1405 bp in length
*      184577 185839: gap of unknown length
*      185839 185939: contig of 1262 bp in length
*      185939 187305: gap of unknown length
*      187305 187404: contig of 1366 bp in length
*      187404 188505: gap of unknown length
*      188505 188605: contig of 1101 bp in length
*      188605 188995: gap of unknown length
*      188995 189896: contig of 1290 bp in length
*      189896 191187: gap of unknown length
*      191187 191287: contig of 1191 bp in length
*      191287 192476: gap of unknown length
*      192476 192577: contig of 1190 bp in length
*      192577 194012: gap of unknown length
*      194012 194111: contig of 1435 bp in length
*      194111 195416: gap of unknown length
*      195416 195515: contig of 1304 bp in length
*      195515 196929: gap of unknown length
*      196929 197029: contig of 1413 bp in length
*      197029 198216: gap of unknown length
*      198216 198316: contig of 1187 bp in length
*      198316 199494: gap of unknown length
*      199494 199594: contig of 1179 bp in length
*      199594 200810: gap of unknown length
*      200810 200910: contig of 1216 bp in length
*      200910 202155: gap of unknown length
*      202155 202255: contig of 1245 bp in length
*      202255 204047: gap of unknown length
*      204047 204147: contig of 1792 bp in length
*      204147 205275: gap of unknown length
*      205275 205376: contig of 1128 bp in length
*      205376 207824: gap of unknown length
*      207824 209566: contig of 2449 bp in length
*      209566 209666: gap of unknown length
*      209666 211262: contig of 1642 bp in length
*      211262 211362: gap of unknown length
*      211362 213638: contig of 1596 bp in length
*      213638 213739: gap of unknown length
*      213739 215671: contig of 2276 bp in length
*      215671 215771: gap of unknown length
*      215771 216917: contig of 1933 bp in length
*      216917 216917: contig of 1146 bp in length
```

\* 216918 217017: gap of unknown length  
 \* 217018 218872: contig of 1855 bp in length.

## FEATURES

source  
 1..218872  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="4"  
 /clone="RP11-210L21"  
 1..2669  
 /note="assembly\_name:Contig100"  
 2670..2769  
 /estimated\_length=unknown  
 2770..4554  
 /note="assembly\_name:Contig101"  
 4555..4654  
 /estimated\_length=unknown  
 4655..6908  
 /note="assembly\_name:Contig102"  
 6909..7008  
 /estimated\_length=unknown  
 7009..10387  
 /note="assembly\_name:Contig103"  
 10388..10487  
 /estimated\_length=unknown  
 10488..12825  
 /note="assembly\_name:Contig104"  
 12826..12925  
 /estimated\_length=unknown  
 12926..15663  
 /note="assembly\_name:Contig105"  
 15664..15763  
 /estimated\_length=unknown  
 15764..18992  
 /note="assembly\_name:Contig106"  
 18993..19092  
 /estimated\_length=unknown  
 19093..21884  
 /note="assembly\_name:Contig107"  
 21885..21984  
 /estimated\_length=unknown  
 21985..25307  
 /note="assembly\_name:Contig108"  
 25308..25407  
 /estimated\_length=unknown  
 25408..38126  
 /note="assembly\_name:Contig109"  
 38127..38226  
 /estimated\_length=unknown  
 38227..67816  
 /note="assembly\_name:Contig110"  
 clone\_end:T7  
 vector\_side:left  
 67817..67916  
 /estimated\_length=unknown  
 67917..175443  
 /note="assembly\_name:Contig111"  
 clone\_end:SP6  
 vector\_side:right  
 175444..175543  
 /estimated\_length=unknown  
 175544..176878  
 /note="assembly\_name:Contig142"  
 176879..176978  
 /estimated\_length=unknown  
 176979..178260  
 /note="assembly\_name:Contig146"  
 178261..178360  
 /estimated\_length=unknown  
 178361..179616  
 /note="assembly\_name:Contig151"  
 179617..179716  
 /estimated\_length=unknown

Query Match 91.6%; Score 17.4; DB 12; Length 218872;  
 Best Local Similarity 94.7%; Pred. No. 3.6e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19

Db 123071 CATTACTTCATAGTCTTTG 123053

## RESULT 82

AC112307/c

## LOCUS

AC112307

AC112307

AC112307.4 GI:24942081

HTG; HTGS PHASE2; HTGS DRAFT; HTGS\_FULLTOP.

Rattus norvegicus (Norway rat)

KEYWORDS

SOURCE

ORGANISM

Rattus norvegicus

Eukaryota; Metazoa;

Mammalia; Eutheria;

Sciurognathi; Muridae;

Murinae; Rattus.

1 (bases 1 to 234431)

REFERENCE

AUTHORS

Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,Y., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Egan,A., Escotto,M., Eugene,C., Evans,C., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebirt,D., Jolivet,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpthy,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulsegad,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmood,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokemele,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L., Pu,L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajds,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villaseana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,R., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,U., Yoon,V., Yu,F., Zhang,J., Zhou,X., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

Direct Submission

TITLE

```

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Unpublished
2 (bases 1 to 234431)
Worley, K.C.
Direct Submission
Submitted (21-FEB-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 234431)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (13-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 13, 2002 this sequence version replaced gi:23603910.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GMYQ
Center clone name: CH230-257P20
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 173236 bases at least Q40
Consensus quality: 175680 bases at least Q30
Consensus quality: 176806 bases at least Q20
Estimated insert size: 179720; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
consists of 1 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces
is believed to be correct as given, however the sizes
of the gaps between them are based on estimates that have
been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
* 234431: contig of 234431 bp in length.
* Location/Qualifiers
  1. .234431
    /organism="Rattus norvegicus"
    /mol_type="genomic DNA"
    /db_xref="taxon:10116"
    /clone="CH230-257P20"
  1. .1221
    /note="wgs_end_extension
clone_end:T7"
46037. .47721
    /note="wgs_end_extension
clone_end:T7"
54017. .54926
    /note="clone_boundary
clone_end:T7"
site:
end sequence:BZ162089"
56917. .57694
    /note="clone_boundary

```

```

clone_end:Sp6
site:
end sequence:BZ162091"
232904. .234431
/note="wgs_end_extension
clone_end:Sp6"

misc_feature
18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Query Match 91.6%; Score 17.4; DB 12; Length 234431;
Best Local Similarity 94.7%; Pred. No. 3.6e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCCTTTG 19
Db 79129 CTTTATTCATAGTCCTTTG 79111

RESULT 83
AC096378/c
LOCUS
DEFINITION
Rattus norvegicus clone CH230-84L6, WORKING DRAFT SEQUENCE.
AC096378
VERSION
AC096378.6 GI:30521763
HTG; HTGS PHASR2; HTGS DRAFT; HTGS_FULLTOP.
KEYWORDS
Rattus norvegicus (Norway rat)
SOURCE
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Rattus.
1 (bases 1 to 235763)
Muzny, D., Marie, Metzker, M., Lee, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Evans, C., Evans, C.A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, B., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorensuewa, L., Loulseghe, H., Lozado, R., Lu, X., Ma, J.,
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martine, E.,
Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwaokeme, O., Okwuonu, G., Olarnpungoon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,
Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,

```



Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission  
Unpublished  
2 (bases 1 to 235763)  
Worley, K.C.

Direct Submission  
Submitted (17-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 235763)  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On May 10, 2003 this sequence version replaced gi:24818898.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information  
Center project name: GFAO  
Center clone name: CH230-84L6  
----- Summary Statistics  
Assembly program: Atlas 3.0;  
Consensus quality: 224184 bases at least Q40  
Consensus quality: 226493 bases at least Q30  
Consensus quality: 227839 bases at least Q20  
Estimated insert size: 235519; sum-of-contigs estimation  
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 1 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\* 1 235763: contig of 235763 bp in length.  
\* Location/Qualifiers  
1..235763  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-84L6"  
1..1226  
/note="wgs contig"  
233871..235763  
/note="wgs\_contig"

FEATURES  
source  
misc\_feature  
misc\_feature  
ORIGIN

Query Match 91.6%; Score 17.4; DB 12; Length 235763;  
Best Local Similarity 94.7%; Pred. No. 3.6e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 CTTTACTTTCATCTCTTTG 19  
|||||  
Db 205307 CTTTACTTTCATGCTTTG 205289

RESULT 84  
AC095186  
LOCUS  
DEFINITION  
Rattus norvegicus clone CH230-9B21, WORKING DRAFT SEQUENCE, 3  
unordered pieces.  
AC095186  
AC095186.6 GI:24941004  
HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
SOURCE  
Rattus norvegicus (Norway rat)  
ORGANISM  
Rattus norvegicus  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Murioidea; Muridae; Murinae; Rattus.  
REFERENCE  
AUTHORS  
Muzny, D., Marie, H., Metzker, M., Lee, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Angiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gabregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guvata, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuhewa, L., Loulseged, H., Lozado, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milobavjevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwakoelameh, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puzo, B., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartabeyn, A., Sisson, I., Sitter, C.D., Smajda, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steelme, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, X., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission

TITLE

```

JOURNAL
REFERENCE 2 (bases 1 to 247690)
AUTHORS Worley, K.C.
TITLE Direct Submission
JOURNAL Submitted (16-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

REFERENCE 3 (bases 1 to 247690)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (13-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

COMMENT
On Nov 13, 2002 this sequence version replaced gi:22777986.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GDBT
Center clone name: CH230-9B21
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 215114 bases at least Q40
Consensus quality: 217886 bases at least Q30
Consensus quality: 219758 bases at least Q20
Estimated insert size: 221396; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 101341: contig of 101341 bp in length
* 101342 101441: gap of unknown length
* 101442 219685: contig of 118244 bp in length
* 219686 219785: gap of unknown length
* 219786 247690: contig of 27905 bp in length.
FEATURES
Location/Qualifiers
source
1..247690
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-9B21"
1..1604
/notes="wgs contig"
misc_feature
101342..101441
gap
/estimated_length=unknown
219686..219785
gap
/estimated_length=unknown
ORIGIN
Query Match 91.6%; Score 17.4; DB 12; Length 247690;
Best Local Similarity 94.7%; Pred. No. 3.6e+02;

```

```

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 CTTTACTTCATAGCTTTTG 19
|||||
Db 200451 CTTTACTTCAAAGCTTTG 200469
|||||

RESULT 85
AC137204/c
LOCUS AC137204
DEFINITION Rattus norvegicus clone CH230-unknown, *** SEQUENCING IN PROGRESS
***, 4 unordered pieces.
AC137204
VERSION AC137204.1 GI:25073092
HTG; HTGS PHAS1; HTGS DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Murioidea; Muridae; Murinae; Rattus.
1 (bases 1 to 249669)
REFERENCE
AUTHORS Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Angiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duvai, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebregorgis, E., Geer, K., Gill, R., Grady, W., Guerra, W., Guevara, W.,
Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, D., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenshehwa, L., Loulsegged, H., Lozado, R. J., Lu, X., Ma, J.,
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwaokemele, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., L.,
Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
Shetty, J., Shvartabeyn, A., Sisson, I., Sitter, C. D., Smajd, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wlaczek, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,
Weinstock, G. and Gibbs, R. A.
Direct Submission
Unpublished
2 (bases 1 to 249669)
TITLE
JOURNAL
REFERENCE

```

**AUTHORS**  
**TITLE** Rat Genome Sequencing Consortium.  
**JOURNAL** Direct Submission  
 Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
**COMMENT** The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: http://www.hgsc.bcm.tmc.edu/  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: KZPB  
 Center clone name: CH230-unknown  
 ----- Summary Statistics  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 201261 bases at least Q40  
 Consensus quality: 206093 bases at least Q30  
 Consensus quality: 208911 bases at least Q20  
 Estimated insert size: 209892; sum-of-contigs estimation  
 Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

-----  
 NOTE: Estimated insert size may differ from sequence length  
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbankdraftdata.html).  
 NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

\* 1 103590: contig of 103590 bp in length  
 \* 103591 103690: gap of unknown length  
 \* 103691 218904: contig of 115214 bp in length  
 \* 218905 219004: gap of unknown length  
 \* 219005 248412: contig of 29408 bp in length  
 \* 248413 248512: gap of unknown length  
 \* 248513 249669: contig of 1157 bp in length.

**FEATURES**  
 source  
 1. 249669  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-unknown"  
 1. 1761  
 /note="wgs contig"  
 4671. 6761  
 /note="wgs contig"  
 103591. 103690  
 /estimated\_length=unknown  
 218905. 219004  
 /estimated\_length=unknown  
 248413. 248512  
 /estimated\_length=unknown

## ORIGIN

Query Match 91.6%; Score 17.4; DB 12; Length 249669;  
 Best Local Similarity 94.7%; Pred. No. 3.6e+02;  
 -Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTACTTCATAGCTTTG 19

Db 123053 CTTACTTCATAGCTTTG 123053

RESULT 86  
 AC108958  
 LOCUS  
 DEFINITION

AC108958 265844 bp DNA linear HTG 13-MAY-2003  
 Rattus norvegicus clone CH230-225123, \*\*\* SEQUENCING IN PROGRESS  
 \*\*\*, 2 unordered pieces.

ACCESSION  
 VERSION  
 HTG; HTGS PHASE1; HTGS DRAFT; HTGS\_ENRICHED.  
 Rattus norvegicus (Norway rat)  
 SOURCE  
 ORGANISM

Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidae; Muridae; Rattus.  
 1 (bases 1 to 265844)

REFERENCE  
 AUTHORS

Muzny, D., Allen, H., Alsebrook, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, T., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensheva, L., Loulseg, H., Lozado, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokemele, O., Okwunu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poinexter, A., Popovic, D., Primus, E., Pu, L.L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scheter, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajd, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlezyk, R., Woodson, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, X., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission

JOURNAL

REFERENCE

2 (bases 1 to 265844)

Worley, K.C.

Direct Submission

JOURNAL

Submitted (03-FEB-2002) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 265844)  
 Rat Genome Sequencing Consortium.  
 Direct Submission  
 Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 On May 13, 2003 this sequence version replaced gi:23321752.  
 The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: http://www.hgsc.bcm.tmc.edu/  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GJEA  
 Center clone name: CH230-225I23  
 ----- Summary Statistics  
 Assembly program: Atlas 3.0;  
 Consensus quality: 241893 bases at least Q40  
 Consensus quality: 247086 bases at least Q30  
 Consensus quality: 249647 bases at least Q20  
 Estimated insert size: 257860; sum-of-contigs estimation  
 Quality coverage: 6x in Q20 bases; sum-of-contigs estimation  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html)  
 \* NOTE: This sequence may represent more than one 'clone'.  
 \* NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved  
 \* 1 262982: contig of 262982 bp in length  
 \* 262983 263082: gap of unknown length  
 \* 263083 265844: contig of 2762 bp in length.  
 Location/Qualifiers  
 misc\_feature  
 1..265844  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-225I23"  
 misc\_feature  
 1..1246  
 /notes="wgs end extension  
 clone\_end:Sp6"  
 misc\_feature  
 1870..3236  
 /notes="wgs end extension  
 clone\_end:Sp6"  
 misc\_feature  
 3572..4405  
 /notes="clone boundary  
 clone\_end:Sp6  
 site:EcORI  
 end sequence:B2101420"  
 misc\_feature  
 complement(257901..258877)  
 /notes="clone boundary  
 clone\_end:T7  
 site:EcORI  
 end sequence:B2101420"

misc\_feature 261286..262982  
 /notes="wgs end extension  
 clone\_end:T7"  
 gap 262983..263082  
 /estimated\_length=unknown  
 ORIGIN  
 Query Match 91.6%; Score 17.4; DB 12; Length 265844;  
 Best Local Similarity 94.7%; Pred. No. 3.6e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 1 CTTTACTTCTACTGCTTTG 19  
 |||||  
 Db 112412 CTTTACTTCTACTGCTTTG 112430  
 RESULT 87  
 AC116211  
 LOCUS  
 DEFINITION Rattus norvegicus clone CH230-69K7, \*\*\* SEQUENCING IN PROGRESS \*\*\*,  
 4 unordered pieces.  
 AC116211  
 VERSION AC116211.4 GI:25007404  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.  
 SOURCE Rattus norvegicus (Norway rat)  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridae; Murinae; Rattus.  
 1 (bases 1 to 268030)  
 Muzny,D., Marie., Metzker,M., Lee., Abramson,S., Adams,C., Alder,J.,  
 Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,  
 Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,  
 Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,  
 Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,  
 Bryant,N., Buay,J., Burch,P., Burrell,K., Calderon,E.,  
 Cardenas,V., Carter,K., Cavazos,I., Caesar,H., Center,A.,  
 Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,  
 Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,  
 Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,  
 Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,  
 Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,  
 Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,I., Fan,G.,  
 Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,  
 Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,  
 Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,  
 Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,  
 Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,  
 Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hughes,M.,  
 Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,  
 Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,  
 Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,  
 Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,  
 Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,  
 Lorensheva,L., Loulseghe,H., Lozado,R.J., Lu,X., Ma,J.,  
 Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,  
 Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,  
 Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,  
 Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,  
 Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,  
 Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,  
 Nwaokemele,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,  
 Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,  
 Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,  
 Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,  
 Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,  
 Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,K., Ruiz,S.J.,  
 Sanders,W., Savary,G., Scherer,S., Scott,G., Shatsman,R., Shen,H.,  
 Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajd,D.,  
 Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,  
 Streimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,  
 Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,  
 Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,

Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, X., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission  
Unpublished  
2 (bases 1 to 268030)  
Worley, K.C.

Direct Submission  
Submitted (26-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 268030)

Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Nov 15, 2002 this sequence version replaced gi:23269678.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu

----- Project Information  
Center project name: GMYB  
Center clone name: CH230-69K7  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 238815 bases at least Q40  
Consensus quality: 241421 bases at least Q30  
Consensus quality: 243001 bases at least Q20  
Estimated insert size: 245676; sum-of-contigs estimation  
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

\* 1 241216: contig of 241216 bp in length  
\* 241217 241316: gap of unknown length  
\* 241317 264502: contig of 23186 bp in length  
\* 264503 264602: gap of unknown length  
\* 264603 266542: contig of 1940 bp in length  
\* 266543 266642: gap of unknown length  
\* 266643 268030: contig of 1388 bp in length.

FEATURES  
source  
Location/Qualifiers  
1..268030  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-69K7"  
1..1142

misc\_feature  
1..1142

-----

Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, X., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission  
Unpublished  
2 (bases 1 to 268030)  
Worley, K.C.

Direct Submission  
Submitted (26-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 268030)

Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Nov 15, 2002 this sequence version replaced gi:23269678.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu

----- Project Information  
Center project name: GMYB  
Center clone name: CH230-69K7  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 238815 bases at least Q40  
Consensus quality: 241421 bases at least Q30  
Consensus quality: 243001 bases at least Q20  
Estimated insert size: 245676; sum-of-contigs estimation  
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

\* 1 241216: contig of 241216 bp in length  
\* 241217 241316: gap of unknown length  
\* 241317 264502: contig of 23186 bp in length  
\* 264503 264602: gap of unknown length  
\* 264603 266542: contig of 1940 bp in length  
\* 266543 266642: gap of unknown length  
\* 266643 268030: contig of 1388 bp in length.

FEATURES  
source  
Location/Qualifiers  
1..268030  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-69K7"  
1..1142

misc\_feature  
1..1142

misc\_feature  
1326..3278  
/note="wgs contig"

gap  
241217..241316  
/estimated\_length=unknown  
264503..264602  
/estimated\_length=unknown  
266543..266642  
/estimated\_length=unknown

ORIGIN  
Query Match 91.6%; Score 17.4; DB 12; Length 268030;  
Best Local Similarity 94.7%; Pred. No. 3.6e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCCTTTG 19  
||||| ||||||| |||||  
Db 22540 CTTTACTTCATAGTCCTTTG 22558

RESULT 88  
AB219542 4343 bp DNA linear BCT 02-JUL-2005  
LOCUS  
DEFINITION  
Campylobacter lari topA, flaA, flaB, CLA0521 genes for  
Topoisomerase I, Flagellins, hypothetical protein CLA0521, partial  
and complete cds, strain: NCTC12892.  
ACCESSION  
AB219542  
VERSION  
AB219542.1 GI:685331170  
SOURCE  
Campylobacter lari  
ORGANISM  
Campylobacter lari  
Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;  
Campylobacteraceae; Campylobacter.

REFERENCE  
1  
AUTHORS  
TITLE  
Variations of two flagellin genes of Campylobacter lari, in  
particular urease-positive thermophilic Campylobacter (UPTC)  
JOURNAL  
Unpublished  
REFERENCE  
2 (bases 1 to 4343)  
AUTHORS  
TITLE  
Submitted (30-JUN-2005) Motoo Matsuda, Azabu University, Laboratory  
of Molecular Biology, School of Environmental Health Sciences;  
1-17-71, Fuchinobe, Kanagawa, 229-8501, Japan  
(E-mail:matsuda@azabu-u.ac.jp, Tel:81-42-769-1942,  
Fax:81-42-754-7661)  
FEATURES  
source  
Location/Qualifiers  
1..4343  
/organism="Campylobacter lari"  
/mol\_type="genomic DNA"  
/strain="NCTC12892"  
/isolation\_source="River water"  
/db\_xref="taxon:201"  
/note="Urease-positive thermophilic Campylobacter"  
gene  
complement(<1..874)  
/gene="topA"  
CDS  
complement(<1..822)  
/gene="topA"  
/codon\_start=1  
/product="Topoisomerase I"  
/protein\_id="BA806130.1"  
/db\_xref="GI:68533171"  
/translation="MKNLIIVESPAAKKTIGNFGKDYEVIAKSHIRDLPKTSFGIK  
IEDENPKPEYRISNDHSLVKELKEKAKAKTIYLADEDEGEAIAYHAKAINOK  
NSLPRIVFHEITKSAIENALNPRSLNNSVNAQOTRRLLDRIYGYKLSPLNOKIOK  
GLSAGRVQSAALKIIVDREIEIKAFVPLKVFSDIMFEKDLQAEVLFQNKIEKLSL  
TNEERAKLIFEACKNANFNKOTESKDRKIAPQAPFTWTSLQOSASRLGNPKKT"  
complement(828..831)  
/gene="topA"  
/note="putative RBS"  
complement(843..848)  
/gene="topA"  
-10\_signal

```
/note="putative prbnow box"
complement(868. .874)
/genes="topA"
/note="putative -35 region"
complement(887. .2477)
/genes="flaA"
complement(887. .918)
/genes="flaA"
/note="putative transcriptional terminator"
complement(936. .2399)
/genes="flaA"
/codon_start=1
/transl_table=11
/product="Flagellin"
/protein_id="BAE06131.1"
/db_xref="GI:68533172"
/translacion="MGPRINTVNASLNAOVNAGNSPALDSSLARLSSGLRINSAAADD
ASGNATADSRNQAASLGQALNNNDIAIGLQTADKAMDEQLKILDTIKVKATQAAQD
GQSAKTRMTIQQEINKLMEELDNIAITTYNGKQLLSGAFNQOQFQVGDKNQNTINAT
IGATQSAKIQTQRTFETGSRITGSGNAGFTIKNDYGVNDFKIQSVILSTSGTGLGALA
AEINKSDKTGVRATATVOTISSGTIOAGNTGDTFTINGVIGKVAVQADKDGSLVA
AINAKDPTTGEASVVGQVLNSADGRGIELSGLTALSGNIASVNYGRLSLVKNDG
SDIILSGSGAGRGTAAREATVNLESVKGOISADIACAMGFNAMSATOPGSKTGCVT
TLQAMVMDIADTAIANDTIIRANIGATONQITSTINNISVTQVNVKAAESQIRDVD
FASANSYKANILAQSGSYAMAQAANAQNVLRLLQ"
/genes="flaA"
complement(2406. .2409)
/note="putative RBS"
complement(2450. .2455)
/genes="flaA"
/note="putative prbnow box"
complement(2472. .2477)
/genes="flaA"
/note="putative -35 region"
2508. .4037
/genes="flaB"
2508. .2509
/genes="flaB"
/note="putative -26 signal"
2520. .2521
/genes="flaB"
/note="putative -13 signal"
2549. .2552
/genes="flaB"
/note="putative RBS"
2559. .4037
/genes="flaB"
/codon_start=1
/transl_table=11
/product="Flagellin"
/protein_id="BAE06132.1"
/db_xref="GI:68533173"
/translacion="MGPRINTGAMNAHANSVITARELDKSLGRLLSGLRINSAAADD
ASGNATADSRNQAASLGQALNNNDIAIGLQTADKAMDEQLKILDTIKVKATQAAQD
GQSAKTRMTIQQEINKLMEELDNIAITTYNGKQLLSGAFNQOQFQVGDKNQNTINAT
IGATQSAKIQTQRTFETGSRITGSGNAGFTIKNDYGVNDFKIQSVILSTSGTGLGALA
AEINKSDKTGVRATATVOTISSGTIOAGNTGDTFTINGVIGKVAVQADKDGSLVA
AINAKDPTTGEASVVGQVLNSADGRGIELSGLTALSGNIASVNYGRLSLVKNDG
SDIILSGSGAGRGTAAREATVNLESVKGOISADIACAMGFNAMSATOPGSKTGCVT
TLQAMVMDIADTAIANDTIIRANIGATONQITSTINNISVTQVNVKAAESQIRDVD
FASANSYKANILAQSGSYAMAQAANAQNVLRLLQ"
complement(4044. .>4343)
/genes="CLA0521"
complement(4044. .>4343)
/genes="CLA0521"
/codon_start=1
/transl_table=11
/product="hypothetical protein CLA0521"
/protein_id="BAE06133.1"
/db_xref="GI:68533174"
/translacion="DLTQDIDKFOILLIKIDNFKVSYNNKQPFYFTEISKPLFYHVE
CKISEIFIKYNDKADKKMKKSVEYLNIETKLWINDYINILKIQNDLLKQTPLYQGV"
4045. .4074
```

```
terminator
/genes="flaB"
4045. .4074
/genes="flaB"
/note="putative transcriptional terminator"

ORIGIN
Query Match 89.5%; Score 17; DB 15; Length 4343;
Best Local Similarity 100.0%; Pred. No. 4.5e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 TTACTTCATAGTCTTTG 19
|||||
Db 744 TTACTTCATAGTCTTTG 760
|||||

RESULT 89
AC172882 67767 bp DNA linear HTG 24-NOV-2005
LOCUS Brassica rapa subsp. pekinensis clone KBrH121H08, *** SEQUENCING IN
DEFINITION PROGRESS ***, 2 ordered pieces.
ACCESSION AC172882
VERSION AC172882.1 GI:82654463
KEYWORDS HTG; HTGS PHASE2
SOURCE Brassica rapa subsp. pekinensis
ORGANISM Brassica rapa subsp. pekinensis
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 67767)
AUTHORS Yang,T.J., Kwon,S.J., Kim,J.A., Park,J.Y., Kim,J.S., Jin,M.,
Lim,K.B., Lim,M.H., Choi,B.S., Lim,Y.P. and Park,B.S.
CONSTRM Korea Brassica Genome Project (KBGP)
TITLE Shotgun sequence of a Brassica rapa BAC clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 67767)
AUTHORS Yang,T.J., Kwon,S.J., Kim,J.A., Park,J.Y., Kim,J.S., Jin,M.,
Lim,K.B., Lim,M.H., Choi,B.S., Lim,Y.P. and Park,B.S.
TITLE Direct Submission
JOURNAL Submitted (24-NOV-2005) Brassica Genomics Team, National Institute
of Agricultural Biotechnology(NIAB), RDA, 224 Seodun-dong, Suwon,
Kyung-Ki Do 441-707, Korea (E-mail:pbeom@rda.go.kr;
Tel:82-31-239-1670; Fax:82-31-299-1672)
COMMENT * NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 12938: contig of 12938 bp in length
* 12939 13038: gap of unknown length
* 13039 67767: contig of 54729 bp in length.
FEATURES Location/Qualifiers
source
1. 67767
/organism="Brassica rapa subsp. pekinensis"
/mol_type="genomic DNA"
/cultivar="Inbred line 'Chiifu'"
/sub_species="pekinensis"
/db_xref="taxon:51351"
/clone="KBrH121H08"
/tissue_type="Young Leaves"
/clone_lib="KBrH (HindIII) BAC Library"
12939..13038
/estimated_length=unknown

gap
12939..13038
/estimated_length=unknown

ORIGIN
Query Match 89.5%; Score 17; DB 12; Length 67767;
Best Local Similarity 100.0%; Pred. No. 5.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TTTACTTCATAGTCTTT 18
```

```
Db      15936 TTTACTTCATAGTCTTT 15952
|||||
RESULT 90
CP000082_03/c
WPCOMMENT
Sequence split into 27 fragments LOCUS CP000082 Accession CP000082
Fragment Name      Begin      End
CP000082_00       1      110000
CP000082_01      100001    210000
CP000082_02      200001    310000
CP000082_03      300001    410000
CP000082_04      400001    510000
CP000082_05      500001    610000
CP000082_06      600001    710000
CP000082_07      700001    810000
CP000082_08      800001    910000
CP000082_09      900001   1010000
CP000082_10     1000001   1100000
CP000082_11     1100001   1210000
CP000082_12     1200001   1310000
CP000082_13     1300001   1400000
CP000082_14     1400001   1510000
CP000082_15     1500001   1610000
CP000082_16     1600001   1710000
CP000082_17     1700001   1810000
CP000082_18     1800001   1910000
CP000082_19     1900001   2010000
CP000082_20     2000001   2110000
CP000082_21     2100001   2210000
CP000082_22     2200001   2310000
CP000082_23     2300001   2410000
CP000082_24     2400001   2510000
CP000082_25     2500001   2610000
CP000082_26     2600001   2650701
Continuation (4 of 27) of CP000082 Psychrobacter arcticus 273

Query Match      89.5%; Score 17; DB 15; Length 110000;
Best Local Similarity 100.0%; Pred. No. 5.6e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      2 TTTACTTCATAGTCTTT 18
|||||
Db      63421 TTTACTTCATAGTCTTT 63405

RESULT 91
AE008384_01
WPCOMMENT
Sequence split into 41 fragments LOCUS AE008384 Accession AE008384
Fragment Name      Begin      End
AE008384_00       1      110000
AE008384_01      100001    210000
AE008384_02      200001    310000
AE008384_03      300001    410000
AE008384_04      400001    510000
AE008384_05      500001    610000
AE008384_06      600001    710000
AE008384_07      700001    810000
AE008384_08      800001    910000
AE008384_09      900001   1010000
AE008384_10     1000001   1110000
AE008384_11     1100001   1210000
AE008384_12     1200001   1310000
AE008384_13     1300001   1410000
AE008384_14     1400001   1510000
AE008384_15     1500001   1610000
AE008384_16     1600001   1710000
AE008384_17     1700001   1810000
AE008384_18     1800001   1910000
AE008384_19     1900001   2010000
AE008384_20     2000001   2110000
```

```
AE008384_21      2100001   2210000
AE008384_22      2200001   2310000
AE008384_23      2300001   2410000
AE008384_24      2400001   2510000
AE008384_25      2500001   2610000
AE008384_26      2600001   2710000
AE008384_27      2700001   2810000
AE008384_28      2800001   2910000
AE008384_29      2900001   3010000
AE008384_30      3000001   3110000
AE008384_31      3100001   3210000
AE008384_32      3200001   3310000
AE008384_33      3300001   3410000
AE008384_34      3400001   3510000
AE008384_35      3500001   3610000
AE008384_36      3600001   3710000
AE008384_37      3700001   3810000
AE008384_38      3800001   3910000
AE008384_39      3900001   4010000
AE008384_40      4000001   4096345
Continuation (2 of 41) of AE008384 from base 100001 (AE008384 Methanosarcina mazei strain

Query Match      89.5%; Score 17; DB 15; Length 110000;
Best Local Similarity 100.0%; Pred. No. 5.6e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      2 TTTACTTCATAGTCTTT 18
|||||
Db      73057 TTTACTTCATAGTCTTT 73073

RESULT 92
AE010299_40
WPCOMMENT
Sequence split into 58 fragments LOCUS AE010299 Accession AE010299
Fragment Name      Begin      End
AE010299_00       1      110000
AE010299_01      100001    210000
AE010299_02      200001    310000
AE010299_03      300001    410000
AE010299_04      400001    510000
AE010299_05      500001    610000
AE010299_06      600001    710000
AE010299_07      700001    810000
AE010299_08      800001    910000
AE010299_09      900001   1010000
AE010299_10     1000001   1110000
AE010299_11     1100001   1210000
AE010299_12     1200001   1310000
AE010299_13     1300001   1410000
AE010299_14     1400001   1510000
AE010299_15     1500001   1610000
AE010299_16     1600001   1710000
AE010299_17     1700001   1810000
AE010299_18     1800001   1910000
AE010299_19     1900001   2010000
AE010299_20     2000001   2110000
AE010299_21     2100001   2210000
AE010299_22     2200001   2310000
AE010299_23     2300001   2410000
AE010299_24     2400001   2510000
AE010299_25     2500001   2610000
AE010299_26     2600001   2710000
AE010299_27     2700001   2810000
AE010299_28     2800001   2910000
AE010299_29     2900001   3010000
AE010299_30     3000001   3110000
AE010299_31     3100001   3210000
AE010299_32     3200001   3310000
AE010299_33     3300001   3410000
AE010299_34     3400001   3510000
AE010299_35     3500001   3610000
AE010299_36     3600001   3710000
```

```

AE010299_37 3700001 3810000
AE010299_38 3800001 3910000
AE010299_39 3900001 4010000
AE010299_40 4000001 4110000
AE010299_41 4100001 4210000
AE010299_42 4200001 4310000
AE010299_43 4300001 4410000
AE010299_44 4400001 4510000
AE010299_45 4500001 4610000
AE010299_46 4600001 4710000
AE010299_47 4700001 4810000
AE010299_48 4800001 4910000
AE010299_49 4900001 5010000
AE010299_50 5000001 5110000
AE010299_51 5100001 5210000
AE010299_52 5200001 5310000
AE010299_53 5300001 5410000
AE010299_54 5400001 5510000
AE010299_55 5500001 5610000
AE010299_56 5600001 5710000
AE010299_57 5700001 5751492
Continuation (41 of 58) of AE010299 from base 4000001 (AE010299 Methanosarcina acetivorans)

Query Match 89.58; Score 17; DB 15; Length 110000;
Best Local Similarity 100.0%; Pred. No. 5.6e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TTTACTTCATAGCTTT 18
Db 76664 TTTACTTCATAGCTTT 76680

RESULT 93
AC155344 110038 bp DNA linear PLN 12-AUG-2005
LOCUS Brassica rapa subsp. pekinensis clone KBrH080A08, complete
DEFINITION
AC155344
AC155344
AC155344.1 GI:57900806
VERSION
KEYWORDS
SOURCE
ORGANISM
Brassica rapa subsp. pekinensis
Brassica rapa subsp. pekinensis
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 110038)
Yang T.J., Kim J.S., Kwon S.J., Kim J.A., Lim K.B., Jin M.,
Park J.Y., Lim M.H., Lim Y.P., Choi B.S. and Park B.S.
KEGG
Complete sequence of a Brassica rapa BAC clone
Unpublished
2 (bases 1 to 110038)
Yang T.J., Kim J.S., Kwon S.J., Kim J.A., Lim K.B., Jin M.,
Park J.Y., Lim M.H., Lim Y.P. and Park B.S.
AUTHORS
Direct Submission
TITLE
Submitted (19-JAN-2005) Brassica Genomics Team, National Institute
of Agricultural Biotechnology (NIAB), RDA, 224 Seodun-dong, Suwon,
Kyung-Ki Do 441-707, Korea (E-mail: pbcom@rda.go.kr;
Tel: 82-31-299-1670; Fax: 82-31-299-1672)
3 (bases 1 to 110038)
Yang T.J., Kim J.S., Kwon S.J., Kim J.A., Lim K.B., Jin M.,
Park J.Y., Lim M.H., Lim Y.P. and Park B.S.
AUTHORS
Direct Submission
TITLE
Submitted (12-AUG-2005) Brassica Genomics Team, National Institute
of Agricultural Biotechnology (NIAB), RDA, 224 Seodun-dong, Suwon,
Kyung-Ki Do 441-707, Korea (E-mail: pbcom@rda.go.kr;
Tel: 82-31-299-1670; Fax: 82-31-299-1672)
This is a finished sequence consisting of 1 contig high quality.
The BAC clone contains flowering locus C gene and located on the
cytogenetic chromosome 10 by genetic map and FISH analyses.
FEATURES
source
1..110038
/organism="Brassica rapa subsp. pekinensis"

```

```

/mol_type="genomic DNA"
/cultivar="Inbred line 'Chiifu'"
/sub_species="pekinensis"
/db_xref="taxon:51351"
/clone="KBrH080A08"
/tissue_type="Young Leaves"
/clone_lib="KBrH (HindIII) BAC Library"
17..2486
/genes="80A08_1"
/notes="Homolog of At5g10020, receptor-like protein kinase
ERECTA, Arabidopsis thaliana, EMBL:AC004484; go_function:
protein serine/threonine kinase activity [goid 0004674];
go_function: ATP binding [goid 0005524]; go_process:
protein amino acid phosphorylation [goid 0006468];
go_process: transmembrane receptor protein tyrosine kinase
signaling pathway [goid 0007169]"
join(17..1256,1332..1909)
/genes="80A08_1"
/notes="Homolog of At5g10020"
/codon_start=1
/product="80A08_1"
/protein_id="AAZ67586.1"
/db_xref="GI:72384470"
/translation="MNSLGGPIPSGRSASSELMAISSDPOMELDLSTNLTGALPGD
IGTMEIRIVLVNANKLSGELSELNKLTSLEFLDLSNDFKQIPEKLPSCWAFNV
SYNDLSGTIPENLRSYPHSFYGNKLRIPGGTPLGSSNGVLAVQAKHDSRLSI
KTAIIVASGAALMLVFLVHRTQKDFHVRNQATTRDAKFGSRSPFLPNTSNA
EQOSSLSFSNDHLTFANGSLSGIFGSEPIFEQALPAASATTINLLDDYPATSGR
KSTSGGSPSSPRFPQVMDLVYSPDLRAGELFFLDVSLKTAELSRAPAEVLGSR
SGHTLYKATLDNGHMLTVKWLVLGVRHKDKPAKAKIKGLKHPNIVPLRAYVWGR
EQERLLLSYMGESLAMLHYETTPRYPSPMSFIQRLKVAVEVAQCLYLHDRAMPHG
LKPNTNIIITSPENTVRIIDYCIHRLMSSGVAEQILNMSALGYSPELASAKPVPT
LKSDVYAFGVILMELLTRRSAGDIISGQGVADLTDMVRLCDQEGRRMDCIDRDIAGG
EEFSKTMEALAIKCIASVNERPIRQVLDLILSLSS"
/genes="80A08_2"
/notes="Homolog of At5g10030, identical to ocs-element
binding factor 4 Gi:414613 from [Arabidopsis thaliana];
go_function: DNA binding [goid 0003677]; go_function:
transcription factor activity [goid 0003700]; go_function:
calmodulin binding [goid 0005516]"
complement(join(5074..5226,5298..5525,5610..5881,
5957..6020,6094..6171,6267..6326,6400..6471,6547..6714))
/genes="80A08_2"
/notes="Homolog of At5g10030"
/codon_start=1
/product="80A08_2"
/protein_id="AAZ67587.1"
/db_xref="GI:72384471"
/translation="MNTTSTHFVPTPRFEIYDPLNQIGTWEEESFKNNGGGFYTPNS
IIPTNQKPYSLSEDEGTGTPHFKDQEASTRHPDKTQRRLAQNREAAKSRURKKAY
VQOLETSLRLKLIHLEQELDRARQQGYFASNRVDTNALSFSDNMCSGIVAFEMEGHW
EQSROISELRTVLNGQVSDIELRLVDNAMKHIFOLFRMKSAALKLDVFIYVMSGMVK
TSAERFLMTGGFRPRLKLLVLPHEPMDMQVLDVCLNRQSCQQAEDAVSQGMELK
QHTLASVAGELGEGSYVPQITSAMERLEALVSFVNQADHLRHETLQOMHRLITTRQ
AARGLALGLEYFORLKRALSSEWETROREP"
10880..11395
/genes="80A08_3"
/notes="Homolog of At5g10040, synonym: T31P16.30; expressed
protein"
10954..11178
/genes="80A08_3"
/notes="Homolog of At5g10040"
/codon_start=1
/product="80A08_3"
/protein_id="AAZ67588.1"
/db_xref="GI:72384472"
/translation="MADIALLVVEEVEERRRLKLTSEKASNSABTGWNDFFARVKNFMS
DQKRWESLKRKLKSAQFALVANGFFSA"
complement(11543..14485)
/genes="80A08_4"
/notes="Homolog of At5g10070, contains Pfam profiles
PF04034: Domain of unknown function (DUF367), PF04068:

```



CDS  
Possible metal-binding domain in RNase L inhibitor, RLI"  
complement(join(11739..12038,12258..12420,12663..12870,  
12974..13089,13183..13254,13346..13436,13527..13663,  
13755..13795,13893..13958,14063..14109,14193..14221,  
14307..14386))  
/gene="80A08\_4"  
/note="Homolog of At5g10070"  
/codon\_start=1  
/product="80A08\_4"  
/protein\_id="AAZ67589.1"  
/db\_xref="GI:72384473"  
/translation="MGYQKRRSRGSGNPNQGTGTENIGREDSLPHDPSGEDEVV  
PNVLAMWDFQCQDKRKLARFLLKELRVNTGGVVLSPVQCSCVSKEDYAL  
IKSLGAVDCSWARLTDVPAKRLPMLVAANPVNPGRCVCEALSA  
LILCEBETANLLGKFKMGHAFLLSKNDILKEYSKCNSAEIISVONSWTQOTLS  
KQAPLKEHVRKEGDEDEDDGLPPLERNMNVISEDEDEDEDEDEDDGLPPL  
ERNVHLEGGDEEDNSGCLCLVSSCGVAEKATASAYQLVVAANSSEEMKK  
SKYKRIIMLKQPPALLTVQGHVIANPAAPQFLKPPVMNPNYFGNIFLMPGPP  
PHMIGNQOPQNPQNSAQGPALQPPGMYYRAPHRGSSGCTILFLIVFLTYPALFH"  
14643..18263  
/gene="80A08\_5"  
/note="Homolog of At5g10080, contains Eukaryotic and viral  
aspartyl proteases active site, PROSIRE:PS00141;  
go\_function: peptin A activity [gold 0004194]; go\_process:  
proteolysis and peptidolysis [gold 0006508]"  
join(14926..15209,15305..15443,15583..15821,15902..16129,  
16210..16419,16495..16588,16680..16743,16843..16914,  
17004..17197,17297..17523,17967..18114)  
/gene="80A08\_5"  
/note="Homolog of At5g10080"  
/codon\_start=1  
/product="80A08\_5"  
/protein\_id="AAZ67590.1"  
/db\_xref="GI:72384474"  
/translation="MASRSAPILLFILSLVSEKSLASLFSSRLIHRFSDEGRASIKSP  
GSFPEKSFYRLLTIDRRQNLGAKFQSLVPSEGSTIKTISFGNYFGWLHTWID  
IGTSPSVFLVALDGSLLWTPCNVOCAPLSSAYYSLSLTKDLNEDFPSASTSKVF  
PCSHKCESAPACSPKEQCPYTVTYASENTSSGLAVEDVLHAYSANASSVVKARV  
VVGCGKEGSEFLGIAPDGMGLGRGEISVPSFLAKGLMRNFSFCFDEEDSGRIY  
FDVGPTQSTQTRFLPNYENFVFGVECCVGNCLQSSFTILDSGQSFTHFLPE  
EYREVALEIDSHINATIKIEGPEWCEYETSPEKVPAILKFSNNFTVTHKPLF  
VLQSEGLVDFCLPISASEGTGGVIGQNYMAGRIYVFDRENKMLGHSASKQCDKIA  
PQAPSPGSTSSNPLPTEQOQTHAVSPAIGKTPKTSASCSFNRLLSSSIL  
LGKAPASTSAHAPRDTCRAGEGTFIFRVKLWFYASHDLGLGNAIAIRVRLGLEI  
ALFAPFLPSLLSLSTHGGAVKVEVTFMAEEDWIVKLDPAPQEKVADSVELLCP  
GLGRCDGG"  
complement(19424..22137)  
/gene="80A08\_6"  
/note="hypothetical protein"  
complement(join(20639..20875,20911..21065,21363..21594))  
/gene="80A08\_6"  
/note="hypothetical protein"  
/codon\_start=1  
/product="80A08\_6"  
/protein\_id="AAZ67591.1"  
/db\_xref="GI:72384475"  
/translation="MQPFLSKLENKPEDRLPLVYNQATQSSQATQSQATQLNPPSPQS  
SQATQSPATQSSASQSPATKSNPADVTSSWGRLLCSARPGTPIYPLVILT  
RVCNVSSTGYLVIIFLNGVKVYINTLPEVETPEFVREEDILDPQVRLSEL  
LHLILALIRQLQRFLLKGDPMYDSNGRYRERDFILVISLTYIS"  
complement(22215..25674)  
/gene="80A08\_7"  
/note="Homolog of At5g10090, contains Pfam profile PF00515  
TPR Domain"  
complement(join(22472..22606,22684..23017,23110..23478,  
23557..24401))  
/gene="80A08\_7"  
/note="Homolog of At5g10090"  
/codon\_start=1  
/product="80A08\_7"  
/protein\_id="AAZ67592.1"  
/db\_xref="GI:72384476"  
/translation="MAENQAEKRSGLLGLGVFGRGRGLWSKKCTADNGNKTTPMRSSNA

gene  
SAPCTSNIQFTKSPGNELNKKLOEYKVSPEPIQNONOTOIQRISKPLSNQYPNNNN  
GPVQOQKVVPRESIGLSESMITDNOKAKMGMLGNLGNLQKPGTTAVGNQTTVO  
NSGGRKTEGERTPTPRTSVSNQOQSGSLCRAISTRMDPEILKINGEDYKNGNF  
VEALADYDAIAIDPKAAAKSNKSAALTALGRIDEAVFECEAIRMEPHYRAHRL  
ANLYLRIGEVANSIYHKGSGPEADQDILKAKTVQHLNCKTEAKRLDWNLLKET  
KNTIAGDAATQVYALQAEAFKSYRHEADDALSCPCFDFVENNTKYGPIGYAGF  
LVVWAQVHMSLGRGEAVEATQLAAKLDNRNREVSMLRRVQAVTAARSKGNDFKGTG  
RFQASAAVGEGLDHDNRNVLNRAACLFKMGQFDRAGIDGSTAALSVRPAYAKARL  
RRADCNAKLGNWELAVGDYELRKETPEDDQVIRGLMEAQHLYKRGHNL"  
28210..32034  
/gene="80A08\_8"  
/note="Homolog of At5g10100, similar to  
trehalose-6-phosphate phosphatase (AtTPP8) [Arabidopsis  
thaliana] GI:2944180; contains Pfam profile PF02358;  
Trehalose-phosphatase; go\_function: trehalose-phosphatase  
Query Match 89.5%; Score 17; DB 4; Length 110038;  
Best Local Similarity 100.0%; Pred. No. 5.6e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 2 TTTACTTCATAGTCTTT 18  
|||||  
Db 7684 TTTACTTCATAGTCTTT 7700  
RESULT 94  
AL359642/c  
LOCUS AL359642 122613 bp DNA linear PRI 18-MAY-2005  
DEFINITION Human DNA sequence from clone RP11-507F17 on chromosome 9 Contains  
a kertain pseudogene, complete sequence.  
ACCESSION AL359642  
VERSION AL359642.12 GI:16972925  
KEYWORDS HTG; keratin.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
1 (bases 1 to 122613)  
REFERENCE  
AUTHORS Dunn, N.  
TITLE Direct Submission  
JOURNAL Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegasanger.ac.uk  
Clome requests: clomequest@sanger.ac.uk  
On Nov 16, 2001 this sequence version replaced gi:15591063.  
The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information  
on the WORMPEP database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping  
Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr9  
RP11-507F17 is from the library RPCI-11.2 constructed by the group  
of Pieter de Jong. For further details see  
http://www.chori.org/bacpac/home.htm  
VECTOR: pBACE3.6  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: vegasanger.ac.uk  
-----  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one subclone; and the assembly was confirmed by restriction digest,  
except on the rare occasion of the clone being a YAC.  
FEATURES:  
Location/Qualifiers  
1..122613  
source

/organism="Homo sapiens"	
/mol_type="genomic DNA"	
/db_xref="taxon:9606"	
/chromosome="9"	
/clone="RP11-507F17"	
/clone_lib="RPC1-11.2"	
/notes="Clone right end: RP11-217F15"	
complement(29127..30282)	
/locus_tag="RP11-507F17.1-001"	
/notes="match: AAH34561 O57607 O93256 P02533	
P02535 P05784 P08728 P08777 P13645 P19001 P19012 Q04695	
Q61414 Q61781"	
/pseudo	
/codon_start=1	
120614	
/notes="Clone_left_end: RP11-165H23"	
89.5%; Score 17; DB 5; Length 122613;	
Best Local Similarity 100.0%; Pred. No. 5.7e+02;	
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY 2 TTTACTTCATAGTCTTT 18	
Db 73999 TTTACTTCATAGTCTTT 73983	
RESULT 95	
AC155845	
LOCUS	
Mus musculus 6 BAC RP24-399D24 (Roswell Park Cancer Institute	
(C57BL/6J Male) Mouse BAC Library) complete sequence.	
ACCESSION	
AC155845 AC115887	
VERSION	
AC155845.3 GI:66793530	
KEYWORDS	
HTG.	
SOURCE	
Mus musculus (house mouse)	
Mus musculus	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;	
Sciurognathi; Muridae; Murinae; Mus.	
1 (bases 1 to 145877)	
Muzny,D., Adams,C., Abai I.I.O., Allen,C., Alsbrooks,S., Archer,P.,	
Arredondo,H., Bandaranaike,D., Bangura,L., Beltran,B., Beltran,R.,	
Beraducci,A., Biswal,K., Blyth,P., Bonham,H., Buhay,C., Burch,P.,	
Cadoree,I., Canada,A., Cardenas,V., Carter,K., Cavazos,I.,	
Chacko,J., Chahrour,M., Chavez,D., Chen,A., Chen,G., Chen,R.,	
Cheng,M.-T., Chu,J., Clerc,K., Cockrell,R., Coyle,M., Cree,A.,	
Curry,S., Dai,W., Davila,M.L., Davis,C., Davy-Carroll,L., De	
Anda,C., Delgado,O., Denson,S., Derramo,C., Ding,Y., Dinh,H.,	
Donlin,J., McCauley,S., Dugan-Rocha,S., Dunn,A., Durbin,K.,	
DiLuca,D., Egan,A., Escotto,M., Espinosa,V., Eugene,C., Fa,M.,	
Fernandez,S., Fernando,P., Flagg,N., Forbes,L., Foster,P.,	
Fowler,G., Fu,Q., Fuh,E., Garcia,A., Garcia,R., Garner,T.,	
Gaskin,C., Gench,S., Ghose,S., Gill,R., Gonzalez,D.,	
Gonzalez-Garay,M., Guevara,W., Holder,M., Haaland,W., Haeblerlen,K.,	
Hall,B., Hamid,H., Hamilton,K., Harbes,B., Harris,R., Havlak,P.,	
Hawes,A., Hawkins,E., Hayes,S., Hemphill,L., Hernandez,J.,	
Hines,S., Hitchens,M., Hodgson,A., Hogue,M., Hollins,B.,	
Howell,L.T., Hulyk,S., Hume,J., Ito,K., Jackson,A., Jackson,L.,	
Jacob,L., Jiang,H., Johnson,B., Johnson,R., Kalatsky,K., Kelly,S.,	
Keys,T., Khan,Z., King,L., Kovar,C., Kowis,A., Kowis,C., Lara,F.,	
Leal,S., Lee,K., Lee,S., LeGall,F.I., Lemon,S., Lewis,L., Li,B.,	
Li,Y., Li,Z., Linnell,M., Liu,W., Liu,Y.-S., Liu,Y., Liyanage,D.,	
London,P., Lopez,J., Lorensuhsuwa,L., Lozada,R., Luk,T., Madu,R.,	
Maheshwari,M., Mahoney,C., Malloy,K., Mansouri,D., Martinez,E.,	
McClelland,H., McPherson,J., Mercadao,C., Metzker,M.,	
Milosavljevic,A., Minja,E., Morgan,M., Morris,S., Munidasa,M.,	
Muray,D., Nazarith,L., Ngo,D., Nguyen,N., Norwig-Eastaugh,E.,	
Nott,A., Nwaokemele,O., Obregon,M., Ochi-Okorie,C., Odeh,E.,	
Okwuonu,G., Okwuonu,K., Parker,D., Pasternak,S., Patel,B.,	
Patei,V., Paul,H., Perez,A., Perez,L., Perrosino,J., Pham,T.,	
Primus,E., Pu,L.-L., Puazo,M., Qin,X., Quinn,A., Quiroz,J.,	
Rabata,D., Rachlin,E., Reigh,R., Ren,Y., Reuter,M., Richards,S.,	
Rives,C., Rodriguez,F., Rojas,A., Ruiz,S.J., Sana,M., Sanders,W.,	
Santibanez,J., Santos,R., Savery,G., Scherer,S., Shen,H., Shen,Y.,	
Sisson,I., Sneed,A., Sodergren,E., Song,X.-Z., Sorrelle,R.,	
Svatek,A., Taylor,E., Taylor,T., Thomas,N., Thorn,R., Thornton,R.,	
Trijos,Z., Usmani,K., Vargo,C., Verduzco,D., Villaseana,D., Virk,D.,	
Volkov,A., Waldron,L., Walker,B., Wang,Q., Wang,S., Warren,J.,	
Wei,X., Wheeler,D., Williams,G., Williams,R., Worley,K., Wright,R.,	
Wu,J., Yakub,S., Yan,K., Yuan,Y., Yu,F., Zhang,J., Zhang,L.,	
Zhang,Z., Zhou,J., Weinstock,G. and Gibbs,R.	
TITLE	
JOURNAL	
REFERENCE	
AUTHORS	
Worley,K.C.	
Direct Submission	
Submitted (20-JAN-2005) Human Genome Sequencing Center, Baylor	
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA	
3 (bases 1 to 145877)	
Direct Submission	
Submitted (27-MAY-2005) Human Genome Sequencing Center, Baylor	
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA	
4 (bases 1 to 145877)	
Worley,K.C.	
Direct Submission	
Submitted (29-MAY-2005) Human Genome Sequencing Center, Baylor	
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA	
On May 29, 2005 this sequence version replaced gi:66773448.	
Sequencing is completed to a minimum standard of double strand	
coverage with a minimum of 2 clones and 2 reads with no ambiguities	
or 2 chemistries with a minimum of 2 clones and 3 reads with no	
ambiguities. If the sequence quality does not meet this standard,	
it will be indicated in the annotation.	
The repeat regions shown were identified using RepeatMasker by	
Adrian Smit.	
Sequence similarities were identified using PowerBlast by Jinghui	
Zhang.	
Exon/Intron boundaries of identified genes were chosen if there	
were canonical splice junctions that maintained sequence continuity	
across the splice junctions.	
----- Genome Center	
Center: Baylor College of Medicine	
Center code: BCM	
Web site: http://www.hgsc.bcm.tmc.edu/	
Drafting Center Code: WIBR	
Contact: hgsc-help@bcm.tmc.edu.	
Location/Qualifiers	
1..145877	
/organism="Mus musculus"	
/mol_type="genomic DNA"	
/db_xref="taxon:10090"	
/chromosome="6"	
/clone="RP24-399D24"	
complement(41..188)	
/rpt_family="B2 Mm2"	
complement(170..263)	
/rpt_family="B1 Mur3"	
complement(211..281)	
/rpt_family="B4"	
282..333	
/rpt_family="AT rich"	
complement(359..1324)	
/rpt_family="RMER178"	
complement(1357..1821)	
/rpt_family="Lx9"	
1823..1842	
/rpt_family="(T)n"	

```
repeat_region complement(1843..1934)
repeat_region /rpt family="B1_Mus2"
repeat_region complement(1934..1998)
repeat_region /rpt family="MTD"
repeat_region complement(2013..2685)
repeat_region /rpt family="Lx4B"
repeat_region complement(2686..2967)
repeat_region /rpt family="Lx5"
repeat_region 3366..3475
repeat_region /rpt family="(TCTA)n"
repeat_region complement(3477..3556)
repeat_region /rpt family="B1_Mus1"
repeat_region 3577..3629
repeat_region /rpt family="(CA)n"
repeat_region 3676..3697
repeat_region /rpt family="AT_rich"
repeat_region 3733..3920
repeat_region /rpt family="B3"
STS 4490..4771
/standard_name="RH125353"
repeat_region 5038..5066
repeat_region /rpt family="(CAAAA)n"
repeat_region 5080..5206
repeat_region /rpt family="B1_Mur1"
repeat_region 5605..5658
repeat_region /rpt family="(CCA)n"
repeat_region 5809..5942
repeat_region /rpt family="B3A"
repeat_region 5966..6129
repeat_region /rpt family="ID_B1"
repeat_region 5996..6130
repeat_region /rpt family="B1_Mur4"
repeat_region 6215..6333
repeat_region /rpt family="RSINE1"
repeat_region complement(6623..6682)
repeat_region /rpt family="B3A"
repeat_region complement(7735..7793)
repeat_region /rpt family="MurSatRep1"
repeat_region 8278..8412
repeat_region /rpt family="B1_Mur4"
repeat_region 8418..8446
repeat_region /rpt family="(A)n"
repeat_region complement(8503..8555)
repeat_region /rpt family="B2_Mm2"
repeat_region complement(8573..8744)
repeat_region /rpt family="B2_Mm1"
repeat_region complement(9439..9550)
repeat_region /rpt family="PB1D9"
repeat_region complement(9577..9696)
repeat_region /rpt family="ID_B1"
repeat_region 9724..9913
repeat_region /rpt family="B3"
repeat_region complement(10256..10404)
repeat_region /rpt family="RSINE1"
repeat_region 10496..10829
repeat_region /rpt family="RMER6A"
repeat_region 11114..11309
repeat_region /rpt family="B3"
repeat_region 11344..11535
repeat_region /rpt family="B2_Mm2"
repeat_region 11686..11751
repeat_region /rpt family="B3A"
repeat_region complement(11821..12010)
repeat_region /rpt family="B3A"
repeat_region complement(12016..12160)
repeat_region 13278..13330
repeat_region /rpt family="AT_rich"
repeat_region 14631..14740
repeat_region /rpt family="B3"
repeat_region 14795..14942
repeat_region /rpt family="B1_Mus1"
repeat_region complement(15333..15395)

/rpt family="RSINE1"
15644..15665
/rpt family="AT_rich"
complement(16347..16538)
/rpt family="B2_Mm1"
16962..17024
/rpt family="ID"
complement(18114..18158)
/rpt family="PB1D9"
complement(18222..18330)
/rpt family="B3A"
complement(18374..18464)
/rpt family="B3A"
complement(18600..18706)

Query Match 89.5%; Score 17; DB 6; Length 145877;
Best Local Similarity 100.0%; Pred. No. 5.7e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TTTACTTCATAGTCTTT 18
|||||
Db 22360 TTTACTTCATAGTCTTT 22376

RESULT 96
CR450699/c
LOCUS CR450699.3 182422 bp DNA linear HTG 12-JAN-2006
DEFINITION Danio rerio chromosome 25 clone DKEIP-120A11, WORKING DRAFT
ACCESSION CR450699
VERSION CR450699.3 GI:84871856
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 182422)
AUTHORS McLaren,S.
TITLE Direct Submission
JOURNAL Submitted (11-JAN-2006) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfish-help@sanger.ac.uk Clone requests:
http://www.sanger.ac.uk/Projects/D_rerio/fags.shtml#dataeight
On Jan 12, 2006 this sequence version replaced gi:67508928.
COMMENT
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk
----- Project Information
Center project name: zkp120A11
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 180066 bases at least Q40
Consensus quality: 180649 bases at least Q30
Consensus quality: 181091 bases at least Q20
Insert size: 181722; sum-of-contigs
Insert size: 189182; 4.3% error; agarose-fp
Quality coverage: 6.23x in Q20 bases; sum-of-contigs Quality
coverage: 6.11x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 11550: contig of 11550 bp in length
* 11551 11650: gap of 100 bp
```

\* 11651 26415: contig of 14765 bp in length  
 \* 26416 26515: gap of 100 bp  
 \* 26516 39653: contig of 13138 bp in length  
 \* 39654 39753: gap of 100 bp  
 \* 39754 42101: contig of 2348 bp in length  
 \* 42102 42201: gap of 100 bp  
 \* 42202 158128: contig of 115927 bp in length  
 \* 158129 158228: gap of 100 bp  
 \* 158229 163016: contig of 4788 bp in length  
 \* 163017 163116: gap of 100 bp  
 \* 163117 166652: contig of 3536 bp in length  
 \* 166653 166752: gap of 100 bp  
 \* 166753 182422: contig of 15670 bp in length.

## FEATURES

source  
 1. .182422  
 /organism="Danio rerio"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:7955"  
 /chromosome="25"  
 /clone\_lib="DanioKeyPilot"  
 /clone\_id="DKEYP-120A11"  
 misc\_feature  
 1. .11550  
 /note="assembly fragment:00170  
 fragment\_chain:1"  
 misc\_feature  
 11651..26415  
 /note="assembly fragment:00086  
 fragment\_chain:1"  
 misc\_feature  
 26516..39653  
 /note="assembly fragment:00415  
 fragment\_chain:1"  
 misc\_feature  
 39754..42101  
 /note="assembly fragment:00011  
 fragment\_chain:2"  
 misc\_feature  
 42202..158128  
 /note="assembly fragment:02027  
 fragment\_chain:2"  
 misc\_feature  
 158229..163016  
 /note="assembly fragment:00018"  
 misc\_feature  
 163117..166652  
 /note="assembly fragment:00046"  
 misc\_feature  
 166753..182422  
 /note="assembly\_fragment:00270.0"

## ORIGIN

Query Match 89.5%; Score 17; DB 12; Length 182422;  
 Best Local Similarity 100.0%; Pred. No. 5.8e+02;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TTTACTTCATAGTCTTT 18  
 |||||  
 Db 18233 TTTACTTCATAGTCTTT 18217

## RESULT 97

AC148203/c  
 LOCUS AC148203 187401 bp DNA linear PRI 09-JUN-2005  
 DEFINITION Callicebus moloch clone LB5-190B16, complete sequence.  
 ACCESSION AC148203  
 VERSION AC148203.3 GI:67078611

## KEYWORDS

HTG

## SOURCE

Callicebus moloch (Dusky titi)  
 Callicebus moloch  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Platyrrhini;  
 Cebidae; Callicebinae; Callicebus.

## REFERENCE

1 (bases 1 to 187401)  
 Antonellis,A., Ayete,K., Bass,D., Benjamin,B., Bera,J.,  
 Blakesley,R.W., Bouffard,G., Brinkley,C., Brooks,S., Chu,G.,  
 Coleman,H., Engle,J., Fukeenko,T., Gestole,M., Greene,A., Guan,X.,  
 Gupta,J., Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Hu,P.,  
 Hunter,G., Hurler,B., Idol,J.R., Kwong,P., Latic,P., Larson,S.,  
 Lee-tin,S.-O., Legaspi,R., Madden,M., Maduro,Q.I., Maduro,V.B.,  
 Margulies,E.H., Masello,C., Maskeri,B., McDowell,J., Mojidi,H.A.,

Mullikin,J.C., Oestreicher,J.S., Park,M., Portnoy,M.E., Prasad,A.,  
 Puri,O., Reddix-Dugue,N., Sante,A., Schandler,K., Schueler,M.G.,  
 Sison,C., Stantripop,S., Stephen,E., Taye,A., Thomas,J.W.,  
 Thomas,P.J., Tsipouri,V., Ung,L., Vogt,J.L., Wetherby,K.D.,  
 Withers,T.R., Young,A. and Green,E.D.

## NISC Comparative Sequencing Initiative

## Unpublished

2 (bases 1 to 187401)

## Green,E.D.

## Direct Submission

Submitted (12-FEB-2004) NIH Intramural Sequencing Center, 8717

Groveton Circle, Gaithersburg, MD 20877, USA

3 (bases 1 to 187401)

## Green,E.D.

## Direct Submission

Submitted (04-MAR-2004) NIH Intramural Sequencing Center, 8717

Groveton Circle, Gaithersburg, MD 20877, USA

4 (bases 1 to 187401)

## Green,E.D.

## Direct Submission

Submitted (09-JUN-2005) NIH Intramural Sequencing Center, 5625

Fishers Lane, Rockville, MD 20852, USA

On Jun 9, 2005 this sequence version replaced gi:44917650.

----- Genome Center

Center: NIH Intramural Sequencing Center

Center code: NISC

Web site: <http://www.nisc.nih.gov>

Contact: [nisc.zoo@nhgri.nih.gov](mailto:nisc.zoo@nhgri.nih.gov)

----- Project Information

Center project name: gte

Center clone name: 190B16

This sequence was finished as follows unless otherwise noted:  
 all regions were double-stranded, sequenced with an  
 alternate chemistry, or covered by high quality data  
 (i.e., phred quality >= 30); an attempt was made to resolve  
 all sequencing problems, such as compressions and repeats;  
 all regions were covered by at least one plasmid subclone  
 or more than one M13 subclone; and the assembly was confirmed  
 by restriction digest.

CLONE LENGTH: This sequence represents the entire insert of  
 this clone unless otherwise noted. If there are overlapping  
 clones, the overlaps are noted in the beginning and end of  
 the Features section.

## FEATURES

## source

1. .187401

/organism="Callicebus moloch"

/mol\_type="genomic DNA"

/db\_xref="taxon:9523"

/clone\_lib="LB5-190B16"

/notes="BAC resource: <http://bacpac.chori.org/>"

48920..48927

/note="single clone coverage"

127297..127302

/note="low quality single stranded/single chemistry

region"

135566..135658

/note="single clone coverage"

## ORIGIN

Query Match 89.5%; Score 17; DB 5; Length 187401;

Best Local Similarity 100.0%; Pred. No. 5.8e+02;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TTTACTTCATAGTCTTT 18

|||||

Db 71590 TTTACTTCATAGTCTTT 71574

## RESULT 98

CR354430

**LOCUS** CR354430 195080 bp DNA linear HTG 08-FEB-2006  
**DEFINITION** Danio rerio chromosome 25 clone CH211-218P2, WORKING DRAFT  
**SEQUENCE**, 2 unordered pieces.  
**ACCESSION** CR354430  
**VERSION** CR354430.13 GI:87080614  
**KEYWORDS** HTG; HTGS PHASE1; HTGS\_ACTIVEFIN; HTGS\_DRAFT; HTGS\_FULLTOP.  
**SOURCE** Danio rerio (zebrafish)  
**ORGANISM** Danio rerio  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
 Cypriniformes; Cyprinidae; Danio.  
**REFERENCE** 1 (bases 1 to 195080)  
**AUTHORS** Heath, P.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (07-FEB-2006) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 zfish-help@sanger.ac.uk Clone requests:  
 http://www.sanger.ac.uk/Projects/D\_rerio/fags.shtml#dataeight  
 On Feb 8, 2006 this sequence version replaced gi:84616971.  
**COMMENT** ----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: <http://www.sanger.ac.uk>  
 Contact: zfish-help@sanger.ac.uk  
 ----- Project Information  
 Center project name: zc218P2  
 ----- Summary Statistics  
 Assembly program: XGAP4; version 4.5  
 Chemistry: Dye-terminator; 100% of reads  
 Consensus quality: 194916 bases at least Q40  
 Consensus quality: 194943 bases at least Q30  
 Consensus quality: 194965 bases at least Q20  
 Insert size: 194980; sum-of-contigs  
 Insert size: 181200; 14.5% error; agarose-fp  
 Quality coverage: 12.17x in Q20 bases; sum-of-contigs Quality  
 coverage: 13.28x in Q20 bases; agarose-fp  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 2 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 106591: contig of 106591 bp in length  
 \* 106592 106691: gap of 100 bp  
 \* 106692 195080: contig of 88389 bp in length.  
**FEATURES** Location/Qualifiers  
 source  
 1..195080  
 /organism="Danio rerio"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:7955"  
 /chromosome="25"  
 /clone="CH211-218P2"  
 /clone\_lib="CHORI-211"  
 1..106591  
 /note="assembly\_fragment:02685"  
 106692..195080  
 /note="assembly\_fragment:04350"  
**misc\_feature**  
**misc\_feature**  
**ORIGIN**  
 Query Match 89.5%; Score 17; DB 12; Length 195080;  
 Best Local Similarity 100.0%; Pred. No. 5.9e+02;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 2 TTTACTTCATAGTCTTT 18  
 |||||  
 Db 88539 TTTACTTCATAGTCTTT 88555  
 RESULT 99  
 AC111133

**LOCUS** AC111133 231443 bp DNA linear HTG 26-MAR-2004  
**DEFINITION** Mus musculus chromosome 6 clone RP23-346C15 map 6, \*\*\* SEQUENCING  
**IN PROGRESS** \*\*\*, 7 unordered pieces.  
**ACCESSION** AC111133  
**VERSION** AC111133.3 GI:45752893  
**KEYWORDS** HTG; HTGS PHASE1; HTGS\_FULLTOP; HTGS\_ACTIVEFIN.  
**SOURCE** Mus musculus (house mouse)  
**ORGANISM** Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridea; Muridae; Murinae; Mus.  
**REFERENCE** 1 (bases 1 to 231443)  
**AUTHORS** Birren, B., Nusbaum, C. and Lander, E.  
**TITLE** Mus musculus chromosome 6, clone RP23-346C15  
**JOURNAL** Unpublished  
**REFERENCE** 2 (bases 1 to 231443)  
**AUTHORS** Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,  
 Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B.,  
 Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,  
 Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,  
 Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,  
 Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,  
 Gande, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
 Hagos, B., Horton, L., Hulme, W., Iliiev, I., Johnson, R., Jones, C.,  
 Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R.,  
 Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C.,  
 Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M.,  
 McEwan, P., McKernan, K., Meldrum, J., Meneus, L., Mihova, T.,  
 Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,  
 Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,  
 Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,  
 Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,  
 Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S.,  
 Sever, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
 Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
 Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,  
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,  
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
**Direct Submission**  
 Submitted (18-FEB-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
**REFERENCE** 3 (bases 1 to 231443)  
**AUTHORS** Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,  
 Anderson, M., Anderson, S., Arachchi, H.M., Barna, N., Bastien, V.,  
 Bloom, T., Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J.,  
 Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B.,  
 DeArellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L.,  
 Erickson, J., Faro, S., Ferreira, P., FitzGerald, M., Gage, D.,  
 Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N.,  
 Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliiev, I.,  
 Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T.,  
 Levine, R., Lindblad-Toh, K., Liu, G., Liu, X., Lui, A., Mabbitt, R.,  
 MacLean, C., Macdonald, P., Major, J., Manning, J., Matthews, C.,  
 McCarthy, M., Meldrum, J., Meneus, L., Mihova, T., Mienga, V.,  
 Murphy, T., Naylor, J., Nguyen, C., Nguyen, T., Nicol, R., Norbu, C.,  
 O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,  
 Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C.,  
 Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schuback, R.,  
 Seaman, S., Sever, P., Smith, C., Spencer, B., Stange-Thomann, N.,  
 Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J.,  
 Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R.,  
 Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L.,  
 Zimmer, A. and Zody, M.  
**Direct Submission**  
 Submitted (26-MAR-2004) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Mar 26, 2004 this sequence version replaced gi:30018044.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
 ----- Genome Center  
 Center: Whitehead Institute/MIT Center for Genome Research  
 Center code: WIBR

Web site: <http://www.seq.wi.mit.edu>  
 Contact: [sequence\\_submissions@road.mit.edu](mailto:sequence_submissions@road.mit.edu)  
 ----- Project Information  
 Center project name: L22671  
 Center clone name: 346\_C\_15

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 7 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 107127: contig of 107127 bp in length  
 \* 107128: gap of 100 bp  
 \* 107228: contig of 13045 bp in length  
 \* 120273: gap of 100 bp  
 \* 120373: gap of 100 bp  
 \* 120378: contig of 25044 bp in length  
 \* 145417: gap of 100 bp  
 \* 145517: contig of 20505 bp in length  
 \* 166022: gap of 100 bp  
 \* 166122: contig of 37245 bp in length  
 \* 203367: gap of 100 bp  
 \* 203467: contig of 4812 bp in length  
 \* 208279: gap of 100 bp  
 \* 208379: contig of 23065 bp in length.

#### FEATURES

source  
 1..231443  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /chromosome="6"  
 /map="6"  
 /clone="RP23-346C15"  
 /clone\_lib="RPCI-23 Female Mouse BAC"  
 107128..107227  
 /estimated\_length=100  
 120273..120372  
 /estimated\_length=100  
 145417..145516  
 /estimated\_length=100  
 166022..166121  
 /estimated\_length=100  
 203367..203466  
 /estimated\_length=100  
 208279..208378  
 /estimated\_length=100

#### ORIGIN

Query Match 89.5%; Score 17; DB 12; Length 231443;  
 Best Local Similarity 100.0%; Pred. No. 5.9e+02;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TTACTTCATAGTCTTT 18  
 |||||

Db 195337 TTTACTTCATAGTCTTT 195353

#### RESULT 100

AC158667  
 LOCUS Mus musculus 6 BAC RP23-83P5 (Roswell Park Cancer Institute  
 DEFINITION (C57Bl/6J Female) Mouse BAC Library) complete sequence.

AC158667

AC158667.8 GI:66792971

KEYWORDS HTG.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 234055)

#### AUTHORS

Muzny, D., Adams, C., Agbai II, O., Allen, C., Alsebrook, S., Archer, P.,  
 Arredondo, H., Bandaranaike, D., Bangura, L., Beltman, B., Beltran, R.,  
 Beraducci, A., Biswal, K., Blyth, P., Bonham, H., Buhay, C., Burch, P.,  
 Cadoree, I., Canada, A., Cardenas, V., Carter, K., Cavazos, I.,  
 Chacko, J., Chahrour, M., Chavez, D., Chen, A., Chen, G., Chen, R.,  
 Cheng, M.-T., Chu, J., Clerc, K., Cockrell, R., Coyle, M., Cree, A.,  
 Curry, S., Dai, W., Davila, M. L., Davis, C., Davy-Carroll, L., De  
 Anda, C., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H.,  
 Donlin, J., McCauley, S., Dugan-Rocha, S., Dunn, A., Durbin, K.,  
 Dziuda, D., Egan, A., Escotto, M., Espinosa, V., Eugene, C., Fa, M.,  
 Fernandez, S., Fernando, P., Flagg, N., Forbes, L., Foster, P.,  
 Fowler, G., Fu, Q., Fuh, E., Garcia, A., Garcia, R., Garner, T.,  
 Gaskin, C., Gench, S., Ghose, S., Gill, R., Gonzalez, D.,  
 Gonzalez-Garay, M., Guevara, W., Holder, M., Haaland, W., Haebler, K.,  
 Hall, B., Hamid, H., Hamilton, K., Harbes, B., Harris, R., Havlak, P.,  
 Hawes, A., Hawkins, E., Hayes, S., Hemphill, L., Hernandez, J.,  
 Hines, S., Hitchens, M., Hodgson, A., Hogues, M., Hollins, B.,  
 Howell, L. T., Hulyk, S., Hume, J., Imo, K., Jackson, A., Jackson, L.,  
 Jacob, L., Jiang, H., Johnson, B., Johnson, R., Kalatus, K., Kelly, S.,  
 Keys, T., Khan, Z., King, L., Kovar, C., Kowis, A., Kowis, C., Lara, P.,  
 Leal, S., Lee, K., Lee, S., Legall, F. I., Lemon, S., Lewis, L., Li, B.,  
 Li, Y., Li, Z., Linnell, M., Liu, W., Liu, Y.-S., Liu, Y., Liyansage, D.,  
 London, P., Lopez, J., Lorensuhewa, L., Lozado, R., Luk, T., Madu, R.,  
 Maheshwari, M., Mahoney, C., Malloy, K., Mansouri, D., Martinez, E.,  
 McElliand, H., McPherson, J., Mercadao, C., Metzker, M.,  
 Milosavljevic, A., Minja, E., Morgan, M., Morris, S., Munidasa, M.,  
 Murray, D., Nazareth, L., Ngo, D., Nguyen, N., Norwig-Eastough, E.,  
 Nott, A., Nwaokemele, O., Obregon, M., Ochi-Okorie, C., Odeh, E.,  
 Okwuonu, G., Okwuonu, K., Parker, D., Pasternak, S., Patel, B.,  
 Patel, V., Paul, H., Perez, A., Perez, L., Petrosino, J., Pham, T.,  
 Primus, E., Pu, L.-L., Puazo, M., Qin, X., Quinn, A., Quroz, J.,  
 Rabata, D., Rachlin, E., Reigh, R., Ren, Y., Reuter, M., Richards, S.,  
 Rives, C., Rodriguez, F., Rojas, A., Ruiz, S. J., Sana, M., Sanders, W.,  
 Santibanez, J., Santos, R., Saverly, G., Scherer, S., Shen, H., Shen, Y.,  
 Sisson, I., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R.,  
 Svatek, A., Taylor, E., Taylor, T., Thomas, N., Thorn, R., Thornton, R.,  
 Trejos, Z., Usmani, K., Vargo, C., Verduzco, D., Villanasa, D., Virk, D.,  
 Volkov, A., Waldron, L., Walker, B., Wang, Q., Wang, S., Warren, J.,  
 Wei, X., Wheeler, D., Williams, G., Williams, R., Worley, K., Wright, R.,  
 Wu, J., Yakub, S., Yan, K., Yuan, Y., Yu, F., Zhang, J., Zhang, L.,  
 Zhang, Z., Zhou, J., Weinstock, G. and Gibbs, R.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 234055)  
 Worley, K.C.  
 Direct Submission  
 Submitted (23-MAR-2005) Human Genome Sequencing Center, Baylor  
 College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 234055)  
 Worley, K.C.  
 Direct Submission  
 Submitted (22-MAY-2005) Human Genome Sequencing Center, Baylor  
 College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 4 (bases 1 to 234055)  
 Worley, K.C.  
 Direct Submission  
 Submitted (28-MAY-2005) Human Genome Sequencing Center, Baylor  
 College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 On May 28, 2005 this sequence version replaced gi:66392614.  
 Sequencing is completed to a minimum standard of double strand  
 coverage with a minimum of 2 clones and 2 reads with no ambiguities  
 or 2 chemistries with a minimum of 2 clones and 3 reads with no  
 ambiguities. If the sequence quality does not meet this standard,  
 it will be indicated in the annotation.

The repeat regions shown were identified using RepeatMasker by  
 Adrian Smit.

Sequence similarities were identified using Powerblast by Jinghui  
 Zhang.

Exon/Intron boundaries of identified genes were chosen if there  
 were canonical splice junctions that maintained sequence continuity

across the splice junctions.

----- Genome Center  
Center: Baylor College Of Medicine  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Drafting Center Code: BCM  
Contact: hgsc-help@bcm.tmc.edu.  
Location/Qualifiers

# FEATURES

## SOURCE

1..234055  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="6"  
/clone="RP23-83P5"

## MISC\_FEATURE

1..26901  
/note="overlaps bases 167483..194383 of clone AC153620"  
/function="clone overlap"

## REPEAT\_REGION

721..745  
/rpt\_family="AT-rich"  
1152..1178  
/rpt\_family="(T)n"  
1628..1682  
/rpt\_family="PB1"  
3165..3250  
/rpt\_family="PB1D7"  
complement(3565..3622)  
/rpt\_family="B3"  
complement(3981..4071)  
/rpt\_family="MIR"  
complement(4431..4563)  
/rpt\_family="MIR"  
complement(6371..6521)  
/rpt\_family="B1\_Mur3"  
complement(6793..6904)  
/rpt\_family="L1MD3"  
7300..7462  
/rpt\_family="B4"  
complement(7838..7938)  
/rpt\_family="PB1D9"  
8635..8701  
/rpt\_family="C-rich"  
complement(8958..9047)  
/rpt\_family="B1\_Mus2"  
9314..9641  
/rpt\_family="ORR1B2"  
9887..10065  
/rpt\_family="B3"  
10224..10308  
/rpt\_family="B1\_Mur3"  
12386..12571  
/rpt\_family="B4"  
12572..12729  
/rpt\_family="B4"  
12899..12942  
/rpt\_family="(CA)n"  
complement(13094..13151)  
/rpt\_family="ID4"  
13333..13552  
/rpt\_family="ID\_B1"  
13616..13814  
/rpt\_family="B3"  
15524..15543  
/rpt\_family="(TTA)n"  
15545..15574  
/rpt\_family="(TG)n"  
complement(15575..15712)  
/rpt\_family="B3"  
complement(15955..16101)  
/rpt\_family="B1\_Mus1"  
complement(16140..16244)  
/rpt\_family="B1\_Mur2"  
complement(16300..16406)  
/rpt\_family="B1F"

repeat\_region complement(16873..16950)  
/rpt\_family="MTD"  
17318..17402  
/rpt\_family="PB1D10"  
17414..17535  
/rpt\_family="B3A"  
complement(17802..17892)  
/rpt\_family="PB1D10"  
20543..20563  
/rpt\_family="AT-rich"  
20844..20958  
/rpt\_family="MLTIF"  
complement(22036..22127)  
/rpt\_family="B4A"  
22355..22409  
/rpt\_family="(CATA)n"  
complement(23280..23412)  
/rpt\_family="B1F"  
23531..23778  
/rpt\_family="MTB"  
24148..24162  
/rpt\_family="A-rich"  
24163..24309  
/rpt\_family="B1\_Mm"  
24310..24355  
/rpt\_family="A-rich"  
24356..24474  
/rpt\_family="B1\_Mus1"  
24479..24530  
/rpt\_family="(CA)n"  
complement(25183..25316)  
/rpt\_family="MER20"  
complement(25556..27435)  
/rpt\_family="L1\_Mus4"  
27483..27628  
/rpt\_family="B1\_Mus1"  
28467..28697  
/rpt\_family="RCHARR1"  
28769..29374  
/rpt\_family="RCHARR1"  
30286..30355  
/rpt\_family="(TG)n"

Query Match 89.5% Score 17; DB 6; Length 234055;  
Best Local Similarity 100.0%; Pred. No. 5.9e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TTTACTTTCATAGTCTTT 18  
|||||

Db 176343 TTTACTTTCATAGTCTTT 176359

Search completed: June 10, 2006, 18:18:21  
Job time : 606.067 secs

**THIS PAGE IS BLANK**



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2006, 15:16:40 ; Search time 141.233 Seconds  
(without alignment)  
937.971 Million cell updates/sec

Title: US-10-600-816-32

Perfect score: 19

Sequence: 1 ctttactcatagctttg 19

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 5244920 seqs, 3486124231 residues

Total number of hits satisfying chosen parameters: 10489840

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

Database :

N Geneseq 8:\*

1: Geneseqn1980s:\*

2: Geneseqn1990s:\*

3: Geneseqn2000s:\*

4: Geneseqn2001as:\*

5: Geneseqn2001bs:\*

6: Geneseqn2002as:\*

7: Geneseqn2002bs:\*

8: Geneseqn2003as:\*

9: Geneseqn2003bs:\*

10: Geneseqn2003cs:\*

11: Geneseqn2003ds:\*

12: Geneseqn2004as:\*

13: Geneseqn2004bs:\*

14: Geneseqn2005s:\*

15: Geneseqn2006s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	19	100.0	19	12	AD128529 Human GPC
2	19	100.0	497	11	ADT96418 Colon can
3	19	100.0	497	11	ADT96418 Colon can
4	19	100.0	552	11	ADT96292 Colon can
5	19	100.0	552	11	ADT96292 Colon can
6	19	100.0	603	3	ADX42774 Human cDN
7	19	100.0	603	3	ADX42774 Human cDN
8	19	100.0	620	6	AZ900046 Hydrophob
9	19	100.0	620	6	AZ900046 Hydrophob
10	19	100.0	634	6	ABQ58527 Human col
11	19	100.0	634	6	ABQ58527 Human col
12	19	100.0	642	10	ABQ59698 Human col
13	19	100.0	1074	12	ABT22328 Breast ca
14	19	100.0	1074	12	ABT22328 Breast ca
15	19	100.0	1212	3	ADO30035 Human GPC
16	19	100.0	1212	3	ADO30035 Human GPC
17	19	100.0	1212	4	AZ61776 cDNA enco
18	19	100.0	1212	4	AAC99709 Skin cell
19	19	100.0	1212	6	ABL34861 Human cDN
20	19	100.0	1212	6	ABL34861 Human cDN
21	19	100.0	1228	8	ACA04775 cDNA enco
22	19	100.0	1228	8	ACA04775 cDNA enco
23	19	100.0	1460	14	AEA00112 Human TAT
24	19	100.0	1460	14	AEA00112 Human TAT
25	19	100.0	1460	14	AEA00632 Human TAT
26	19	100.0	1619	4	AAF58615 Human REC
27	19	100.0	1619	4	AAF58615 Human REC
28	19	100.0	1718	3	AAZ90056 Hydrophob
29	19	100.0	1718	3	AAZ90056 Hydrophob

Abt10173 Human bre	2302	6	ABT10173
Acc58386 Human GPC	2302	8	ACC58386
Add93240 RAIG1 cod	2302	10	ADD93240
Adl24773 Intestina	2302	10	ADL24773
Abt31923 Human bre	2316	4	ABT31923
Aah14688 Human cDN	2446	4	AAH14688
Abz242832 Human G p	2456	8	ABZ242832
Acc72695 Human can	2456	10	ACC72695
Adn39301 Cancer/an	2456	11	ADN39301
Adi28525 Human GPC	2456	12	ADI28525
Adi28459 Human GPC	2456	12	ADI28459
Adr48221 Human ret	2456	13	ADR48221
Acn39388 Tumour-as	2456	13	ACN39388
Adr43992 Human bre	2456	13	ADR43992
Adu06126 Novel bro	2456	13	ADU06126
Rea00088 Human TAT	2456	14	REA00088
Aea00608 Human TAT	2456	14	AEA00608
Aed47480 Retinoloic	2456	14	AED47480
Abq54954 Human ova	2593	6	ABQ54954
Aah72766 Human cer	3371	4	AAH72766
Acn89274 Breast ca	4239	11	ACN89274
Adx97494 Pancreatit	6730	13	ADX97494
Aaa96701 Reporter	1725	3	AAA96701
Adf70573 Orphan re	1788	10	ADF70573
Aef80156 Cancer-as	18273	15	AEF80156
Aee07706 Tobacco p	168	14	AEE07706
Aef98099 Tobacco c	168	15	AEF98099
Aaa62561 Human bre	354	3	AAA62561
Abt12676 Orestes s	354	6	ABT12676
Acn91970 Human col	354	10	ACN91970
Aac28686 Human sec	407	3	AAC28686
Aak74755 Human imm	429	4	AAK74755
Aak88468 Human dig	491	4	AAK88468
Aai97742 Human neu	910	4	AAI97742
Aai97741 Human neu	981	4	AAI97741
Aak90456 Human dig	1596	4	AAK90456
Aak90457 Human dig	1596	4	AAK90457
Aef99265 Human EFH	3447	15	AEF99265
Adj80212 Novel hum	3737	10	ADJ80212
Aec20314 Human nuc	3737	14	AEC20314
Aac75928 Human ORF	3739	3	AAC75928
Acn39480 Tumour-as	3786	15	ACN39480
Aea61137 Human BCA	4816	13	AEA61137
Aac36870 Arabidops	127567	14	AAC36870
Adk16126 Nanoarcha	483	3	ADK16126
Adn73034 Thale cre	654	12	ADN73034
Aas25932 Human cDN	1131	12	AAS25932
Adb1647 Human ova	1199	4	ADB1647
Abx73273 Human nov	1258	10	ABX73273
Acc43801 Nucleotid	1259	8	ACC43801
Aac49655 Arabidops	1366	8	AAC49655
Aac33429 Arabidops	1394	3	AAC33429
Aai60916 Human pol	1398	3	AAI60916
Adb83252 Human cDN	1401	4	ADB83252
Asa64753 DNA encod	1599	9	ASA64753
Adn38895 Cancer/an	1734	5	ADN38895
Acc43802 Nucleotid	1734	11	ACC43802
Adj41386 Plant cDN	1868	8	ADJ41386
Ab160217 Human enc	2000	12	AB160217
Aah16484 Human cDN	2021	4	AAH16484
Ab120545 Drosophil	2063	4	ABL20545
Abq72580 Human MDD	2094	4	ABQ72580
Adz26902 Adeno-ass	2165	4	ADZ26902
Adz26918 Adeno-ass	2175	6	ADZ26918
Adz26920 Adeno-ass	2208	14	ADZ26920
Adz26903 Adeno-ass	2208	14	ADZ26903
Adz26905 Adeno-ass	2208	14	ADZ26905
Adz26906 Adeno-ass	2208	14	ADZ26906
Adz26900 Adeno-ass	2208	14	ADZ26900













c 968 14.8 77.9 65558 14 AEA61193  
969 14.8 77.9 66126 13 ABD33341  
970 14.8 77.9 72821 13 ABD32900  
971 14.8 77.9 81679 12 ADQ59189  
972 14.8 77.9 82660 11 ACN45192  
973 14.8 77.9 82952 6 ABN5766  
974 14.8 77.9 83400 12 ADP07906  
975 14.8 77.9 90885 12 ADX16049\_4  
976 14.8 77.9 92407 4 AAF28549  
977 14.8 77.9 94330 11 ACN44662  
978 14.8 77.9 96593 9 ADA02750  
979 14.8 77.9 96593 10 ADB72488  
980 14.8 77.9 96593 10 ADC85230  
981 14.8 77.9 96593 12 ADM74345  
982 14.8 77.9 96595 9 ADA02936  
983 14.8 77.9 96595 10 ADB72674  
984 14.8 77.9 96595 10 ADC85416  
985 14.8 77.9 96595 12 ADM74531  
986 14.8 77.9 96596 10 ADC85287  
987 14.8 77.9 96597 9 ADA02807  
988 14.8 77.9 96597 10 ADB72545  
989 14.8 77.9 96597 12 ADM74402  
990 14.8 77.9 96599 9 ADA02831  
991 14.8 77.9 96599 10 ADB72569  
992 14.8 77.9 96599 10 ADC85310  
993 14.8 77.9 96599 12 ADM74426  
994 14.8 77.9 99916 6 ADI03931  
995 14.8 77.9 101505 11 ACN44694  
996 14.8 77.9 101786 3 AAF22293  
997 14.8 77.9 110000 2 AAT42063\_17  
998 14.8 77.9 110000 2 AAV21209\_11  
999 14.8 77.9 110000 2 AAX20248\_01  
1000 14.8 77.9 110000 2 AAZ01425\_01

## ALIGNMENTS

RESULT 1  
ADI28529  
ID ADI28529 standard; DNA; 19 BP.  
XX AC  
AC ADI28529;  
XX AC  
XX DT 22-APR-2004 (first entry)  
XX DE Human GPCR retinoic acid induced 3 (RAI-3) peptide DNA primer.  
XX DE  
XX DE  
XX DE Retinoic acid induced 3; RAI-3; human; G-protein coupled receptor; GPCR;  
KW antiinflammatory; immunosuppressive; cytostatic; cardiant; antiallergic;  
KW broncholytic; gene therapy; PCR; primer; ss.  
XX  
XX Homo sapiens.  
XX OS  
XX WO2004001060-A2.  
XX PN  
XX PD 31-DEC-2003.  
XX PF 20-JUN-2003; 2003WO-US019255.  
XX PR 20-JUN-2002; 2002US-0390850P.  
XX PR 29-AUG-2002; 2002US-0407006P.  
XX PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
XX PI Whitney GS, Opitck G, Garulacan L, Ramanathan CS, McKinnon M;  
PI Bennett KL, Barber LE, Cacace A, Tsuchihashi Z;  
XX DR WPI; 2004-090973/09.  
XX PT New nucleic acid molecule encoding a human G-protein coupled receptor.  
PT RAI3, useful for preventing, treating or ameliorating chronic obstructive  
PT pulmonary disease (COPD), COPD-like disorder, or the underlying symptoms

PT of COPD.  
XX Disclosure; SEQ ID NO 32; 301pp; English.  
PS The present sequence is that of an antisense PCR primer based on DNA  
XX encoding a peptide ADI28458 from human G-protein coupled receptor  
CC retinoic acid induced 3 (RAI-3) ADI28460. The peptide was identified in  
CC studies of RAI-3 phosphorylation in response to cigarette smoke. The  
CC primer can be used in a PCR to identify individuals at risk for  
CC developing chronic obstructive pulmonary disease (COPD). RAI-3 is  
CC tyrosine phosphorylated and/or is associated/complexed with tyrosine  
CC phosphorylated proteins only in cells that have been exposed to cigarette  
CC smoke. Since RAI-3 is primarily expressed in lung tissue, and since a  
CC cigarette smoke is a major causative factor of COPD, RAI-3 provides a  
CC novel cellular target for identifying modulators, e.g. agonists or  
CC antagonists, useful for the treatment and/or prevention of COPD and  
CC related disorders such as emphysema and chronic bronchitis. RAI-3  
CC modulators, e.g. agonists and antagonists, especially antisense  
CC compounds, can be used to treat COPD and other disorders an diseases  
CC associated with regulation of NF-kB and/or its associated or interacting  
CC signaling molecules. Single nucleotide polymorphisms (SNPs) detected in  
CC the RAI-3 gene are useful for determining COPD association in  
CC individuals. RAI-3 nucleic acid molecules and polypeptides are useful for  
CC preventing, treating or ameliorating disorders related to aberrant GPCR  
CC signaling or cell cycle regulation, pulmonary disorders, inflammatory  
CC lung disorders, COPD, the underlying symptoms of COPD, COPD-related  
CC disorders or conditions, autoimmune disorders, disorders related to  
CC hyperimmune activity, inflammatory conditions, disorders related to  
CC aberrant acute phase responses, hypercongenital conditions, birth  
CC defects, necrotic lesions, wounds, organ transplant rejection, renal  
CC diseases, ischaemia-reperfusion injury, heart disorders, disorders  
CC related to aberrant signal transduction, proliferation disorders,  
CC cancers, HIV infection, asthma, cystic fibrosis, pulmonary fibrosis,  
CC ulcerative colitis, cerebral infarct, myocardial infarct, diabetic  
CC nephropathy, allergic rhinitis, Crohn's disease, atherosclerosis,  
CC rheumatoid arthritis, inflammatory/autoimmune disorders, glioblastoma,  
CC pulmonary small cell undifferentiated carcinoma, carcinoma of the breast,  
CC colon, lung, ovary, pancreas, prostate, non-Hodgkin's lymphoma, disorders  
CC associated with aberrant cell adhesion, I-CAM function and/or regulation,  
CC E-selectin function and/or regulation, or aberrant NF-kB function and/or  
XX regulation (all claimed).  
XX SQ Sequence 19 BP; 3 A; 4 C; 2 G; 10 T; 0 U; 0 Other;  
Query Match 100.0%; Score 19; DB 12; Length 19;  
Best Local Similarity 100.0%; Pred. NO. 40;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 CTTTACTTCATAGCTTTG 19  
Db 1 CTTTACTTCATAGCTTTG 19  
RESULT 2  
ADT96418/c  
ID ADT96418 standard; cDNA; 497 BP.  
XX AC  
AC ADT96418;  
XX AC  
XX DT 16-DEC-2004 (first entry)  
XX DE Colon cancer associated human cDNA sequence #1925.  
XX DE  
XX DE Colon cancer; T cell; tumour protein; C634S; C635S; C637S; C640S; C636S;  
KW humoral immune response; cellular immune response; cytostatic;  
KW immunostimulant; human; ss.  
XX OS  
XX Homo sapiens.  
XX PN US2003087818-A1.  
XX PD 08-MAY-2003.



```

PF 01-FEB-2002; 2002US-0006543.
XX
XX 02-FEB-2001; 2001US-0267400P.
XX
PR 07-FEB-2001; 2001US-0267382P.
PR
PR 11-MAY-2001; 2001US-0290322P.
PR
PR 12-JUL-2001; 2001US-0305265P.
PR
PR 16-AUG-2001; 2001US-0313077P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Jiang Y, Chenault RA, Xu J, Indirias CY, Lodes MJ, Secrist H;
XX Carter D, Fanger GR, Smith CL, Durham M, Stolk JA;
XX WPI; 2003-040540/03.
XX
XX New isolated nucleic acids and polypeptides capable of eliciting a
XX humoral and/or cellular immune response, useful for diagnosing,
XX preventing or treating cancer, particularly colon cancer.
XX
XX Claim 1; SEQ ID NO 1937; 87pp; English.
XX
XX The invention relates to polynucleotide and polypeptide sequences
XX associated with cancer, particularly colon cancer. Also disclosed are (i)
XX an expression vector comprising the polynucleotide, (ii) a host cell
XX transformed or transfected with the expression vector, (iii) an isolated
XX antibody, or its antigen-binding fragment, which specifically binds to
XX the polypeptide, (iv) a method of detecting or determining the presence
XX of cancer in a patient, (v) a fusion protein comprising at least one of
XX the polypeptides, (vi) an oligonucleotide that hybridises to the
XX polynucleotide sequence under highly stringent conditions, and (vii) a
XX method of stimulating and/or expanding T cells specific for a tumour
XX protein. The polypeptide specifically comprises the amino acid sequence
XX of C634S, C635S, C637S, C640S, C636S or one of the potential open reading
XX frames (ORFs) of C636S. These polypeptides are encoded by the
XX polynucleotide sequences, where both are capable of eliciting a humoral
XX and/or cellular immune response. The polynucleotides, polypeptides, and
XX antibodies are useful for diagnosing, preventing or treating cancer,
XX particularly colon cancer. The polynucleotide and polypeptide sequences
XX are also useful in DNA strand invasion, antisense inhibition, mutational
XX analysis, nucleic acid purification, isolation of transcriptionally
XX active genes, blocking or transcription factor binding, genome cleavage
XX or in situ hybridisation, and as enhancers of transcription or
XX biomarkers. This sequence represents a human colon cancer associated
XX cDNA. Note: The sequence data for this patent was obtained in electronic
XX format directly from the USPTO web site at seqdata.uspto.gov
XX
XX Sequence 497 BP; 126 A; 130 C; 129 G; 112 T; 0 U; 0 Other;
XX
XX Query Match 100.0%; Score 19; DB 11; Length 497;
XX Best Local Similarity 100.0%; Pred. No. 50;
XX Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 1 CTTTACTTCATAGTCTTTG 19
XX |||||
XX Db 385 CTTTACTTCATAGTCTTTG 367
XX
XX RESULT 4
XX ADT96292
XX ID ADT96292 standard; cDNA; 552 BP.
XX
XX AC ADT96292;
XX
XX 16-DEC-2004 (first entry)
XX
XX Colon cancer associated human cDNA sequence #1799.
XX
XX Colon cancer; T cell; tumour protein; C634S; C635S; C637S; C640S; C636S;
XX humoral immune response; cellular immune response; cytostatic;
XX immunostimulant; human; ss.
XX
XX Homo sapiens.
XX
XX US2003087818-A1.
XX
XX 08-MAY-2003.
XX
XX 01-FEB-2002; 2002US-00066543.
XX
XX 02-FEB-2001; 2001US-0267400P.
XX
XX 07-FEB-2001; 2001US-0267382P.
XX
XX 11-MAY-2001; 2001US-0290322P.
XX
XX 12-JUL-2001; 2001US-0305265P.
XX
XX 16-AUG-2001; 2001US-0313077P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Jiang Y, Chenault RA, Xu J, Indirias CY, Lodes MJ, Secrist H;
XX Carter D, Fanger GR, Smith CL, Durham M, Stolk JA;
XX WPI; 2003-040540/03.
XX
XX New isolated nucleic acids and polypeptides capable of eliciting a
XX humoral and/or cellular immune response, useful for diagnosing,
XX preventing or treating cancer, particularly colon cancer.
XX
XX Claim 1; SEQ ID NO 1937; 87pp; English.
XX
XX The invention relates to polynucleotide and polypeptide sequences
XX associated with cancer, particularly colon cancer. Also disclosed are (i)
XX an expression vector comprising the polynucleotide, (ii) a host cell
XX transformed or transfected with the expression vector, (iii) an isolated
XX antibody, or its antigen-binding fragment, which specifically binds to
XX the polypeptide, (iv) a method of detecting or determining the presence
XX of cancer in a patient, (v) a fusion protein comprising at least one of
XX the polypeptides, (vi) an oligonucleotide that hybridises to the
XX polynucleotide sequence under highly stringent conditions, and (vii) a
XX method of stimulating and/or expanding T cells specific for a tumour
XX protein. The polypeptide specifically comprises the amino acid sequence
XX of C634S, C635S, C637S, C640S, C636S or one of the potential open reading
XX frames (ORFs) of C636S. These polypeptides are encoded by the
XX polynucleotide sequences, where both are capable of eliciting a humoral
XX and/or cellular immune response. The polynucleotides, polypeptides, and
XX antibodies are useful for diagnosing, preventing or treating cancer,
XX particularly colon cancer. The polynucleotide and polypeptide sequences
XX are also useful in DNA strand invasion, antisense inhibition, mutational
XX analysis, nucleic acid purification, isolation of transcriptionally
XX active genes, blocking or transcription factor binding, genome cleavage
XX or in situ hybridisation, and as enhancers of transcription or
XX biomarkers. This sequence represents a human colon cancer associated
XX cDNA. Note: The sequence data for this patent was obtained in electronic
XX format directly from the USPTO web site at seqdata.uspto.gov
XX
XX Sequence 497 BP; 126 A; 130 C; 129 G; 112 T; 0 U; 0 Other;
XX
XX Query Match 100.0%; Score 19; DB 11; Length 497;
XX Best Local Similarity 100.0%; Pred. No. 50;
XX Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 1 CTTTACTTCATAGTCTTTG 19
XX |||||
XX Db 385 CTTTACTTCATAGTCTTTG 367
XX
XX RESULT 3
XX ADX42900/c
XX ID ADX42900 standard; cDNA; 497 BP.
XX
XX AC ADX42900;
XX
XX 21-APR-2005 (first entry)
XX
XX Human cDNA encoding colon cancer protein SEQ ID NO 1937.
XX
XX Cytostatic; Immunostimulant; therapy; diagnosis; colon cancer; neoplasm;
XX ss; gene.
XX
XX Homo sapiens.
XX
XX WO200274156-A2.
XX
XX PN

```

```
XX PA (CORI-) CORIXA CORP.
XX PI Jiang Y, Chenault RA, Xu J, Indirias CV, Lodes MJ, Secretist H;
XX PI Carter D, Fanger GR, Smith CL, Durham M, Stolk JA;
XX DR WPI; 2003-040540/03.
XX
XX PT New isolated nucleic acids and polypeptides capable of eliciting a
XX PT humoral and/or cellular immune response, useful for diagnosing,
XX PT preventing or treating cancer, particularly colon cancer.
XX PS Claim 1; SEQ ID NO 1811; 87pp; English.
XX
XX CC The invention relates to polynucleotide and polypeptide sequences
XX CC associated with cancer, particularly colon cancer. Also disclosed are (i)
XX CC an expression vector comprising the polynucleotide, (ii) a host cell
XX CC transformed or transfected with the expression vector, (iii) an isolated
XX CC antibody, or its antigen-binding fragment, which specifically binds to
XX CC the polypeptide, (iv) a method of detecting or determining the presence
XX CC of cancer in a patient, (v) a fusion protein comprising at least one of
XX CC the polypeptides, (vi) an oligonucleotide that hybridizes to the
XX CC polynucleotide sequence under highly stringent conditions, and (vii) a
XX CC method of stimulating and/or expanding T cells specific for a tumour
XX CC protein. The polypeptide specifically comprises the amino acid sequence
XX CC of C634S, C635S, C637S, C640S, C636S or one of the potential open reading
XX CC frames (ORFs) of C636S. These polypeptides are encoded by the
XX CC polynucleotide sequences, where both are capable of eliciting a humoral
XX CC and/or cellular immune response. The polynucleotides, polypeptides, and
XX CC antibodies are useful for diagnosing, preventing or treating cancer,
XX CC particularly colon cancer. The polynucleotide and polypeptide sequences
XX CC are also useful in DNA strand invasion, antisense inhibition, mutational
XX CC analysis, nucleic acid purification, isolation of transcriptionally
XX CC active genes, blocking or transcription factor binding, genome cleavage
XX CC or in situ hybridisation, and as enhancers of transcription or
XX CC biomarkers. This sequence represents a human colon cancer associated
XX CC cDNA. Note: The sequence data for this patent was obtained in electronic
XX CC format directly from the USPTO web site at seqdata.uspto.gov
XX
XX SQ Sequence 552 BP; 126 A; 141 C; 148 G; 137 T; 0 U; 0 Other;

Query Match 100.0%; Score 19; DB 11; Length 552;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGCTCTTG 19
Db 113 CTTTACTTCATAGCTCTTG 131

RESULT 5
ADX42774
ID ADX42774 standard; cDNA; 552 BP.
XX AC ADX42774;
XX
XX DT 21-APR-2005 (first entry)
XX
XX DE Human cDNA encoding colon cancer protein SEQ ID NO 1811.
XX
XX KW Cytostatic; Immunostimulant; therapy; diagnosis; colon cancer; neoplasm;
XX KW ss; gene.
XX
XX OS Homo sapiens.
XX
XX PN WO200274156-A2.
XX
XX PD 26-SEP-2002.
XX
XX PX 01-FEB-2002; 2002WO-US002870.
XX
XX PR 02-FEB-2001; 2001US-0267400P.
XX PR 07-FEB-2001; 2001US-0267382P.

11-MAY-2001; 2001US-0290322P.
12-JUL-2001; 2001US-0305265P.
16-AUG-2001; 2001US-0313077P.
(CORI-) CORIXA CORP.
Jiang Y, Chenault RA, Xu J, Indirias CV, Lodes MJ, Secretist H;
Carter D, Fanger GR, Smith CL, Durham M, Stolk JA;
WPI; 2003-040540/03.
New isolated nucleic acids and polypeptides capable of eliciting a
humoral and/or cellular immune response, useful for diagnosing,
preventing or treating cancer, particularly colon cancer.
Claim 1; SEQ ID NO 1811; 244pp; English.
The invention relates to a new isolated nucleic acid. The nucleic acids,
polypeptides, antibodies are useful for diagnosing, preventing or
treating cancer, particularly colon cancer. The nucleic acid and
polypeptides are also useful in DNA strand invasion, antisense
inhibition, mutational analysis, nucleic acid purification, isolation of
transcriptionally active genes, blocking or transcription factor binding,
genome cleavage or in situ hybridization, and as enhancers of
transcription or biomarkers. The kits are useful for detecting antibody
binding. The present sequence represents a human cDNA encoding a colon
cancer protein.
Sequence 552 BP; 126 A; 141 C; 148 G; 137 T; 0 U; 0 Other;

Query Match 100.0%; Score 19; DB 11; Length 552;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGCTCTTG 19
Db 113 CTTTACTTCATAGCTCTTG 131

RESULT 6
AAZ90046/c
ID AAZ90046 standard; cDNA; 603 BP.
XX AC AAZ90046;
XX
XX DT 09-MAY-2000 (first entry)
XX
XX DE Hydrophobic domain containing protein clone HP10549 coding sequence.
XX
XX KW Hydrophobic domain; clone HP10549; nutritional supplement; SCID; HIV;
XX KW cell proliferation; immune stimulant; immune deficiency; tumour; pain;
XX KW rheumatoid arthritis; insulin dependent diabetes mellitus; fertility;
XX KW myasthenia gravis; haematopoiesis regulator; tissue growth; depression;
XX KW anti-inflammatory; infection; bodily characteristic; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200000506-A2.
XX
XX PD 06-JAN-2000.
XX
XX PF 18-JUN-1999; 99WO-JP003242.
XX
XX PR 26-JUN-1998; 98JP-00180008.
XX
XX PA (SAGA ) SAGAMI CHEM RES CENT.
XX PA (PROT-) PROTEGENE INC.
XX
XX PI Kato S, Kimura T;
XX
XX DR WPI; 2000-160665/14.
XX DR P-PSDB; AAY78809.
```

PT Novel human proteins having hydrophobic domains used for research and  
PT diagnostic purposes.  
XX  
PS Claim 3; Page 88; 117pp; English.  
XX  
CC This sequence represents the hydrophobic domain containing protein, clone  
CC HP10549 coding region. The sequence is isolated from a human stomach  
CC cancer cell line. The invention relates to human proteins with  
CC hydrophobic domains, the DNA and the cDNA encoding them. The  
CC polynucleotides and proteins are predicted to have biological activities  
CC which make them suitable for treating, preventing or ameliorating medical  
CC conditions in humans and animals. Suggested activities include  
CC nutritional activity (nutritional source or supplement); cytokine and  
CC cell proliferation/differentiation activity; immune stimulating (e.g. as  
CC vaccines) or suppressing activity (e.g. to treat various immune  
CC deficiencies such as SCIDS or HIV, connective tissue disease, systemic  
CC lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary  
CC inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin  
CC dependent diabetes mellitus, myasthenia gravis, graft-versus-host disease  
CC and autoimmune inflammatory eye disease, as well as asthma, allergies and  
CC organ transplantation); haematopoiesis regulating activity (e.g. in  
CC treatment of myeloid or lymphoid cell deficiencies); tissue growth  
CC activity (e.g. wound healing and tissue repair, ulcers, burns,  
CC periodontal disease); activin/inhibin activity; chemotactic/chemokinetic  
CC activity; haemostatic and thrombolytic activity (e.g. treating  
CC haemophilias); receptor/ligand activity; anti-inflammatory activity; and  
CC tumour inhibition activity. The polynucleotides are also stated to be  
CC useful for gene therapy. Other activities include inhibiting infections  
CC caused by bacteria, fungi, viruses and other parasites (e.g. Hepatitis,  
CC malaria); effecting bodily characteristics such as, e.g. weight, colour,  
CC skin, effecting biorhythms or cardiac cycles; enhancing fertility;  
CC treatment of depression; treatment of pain; hormonal or endocrine  
CC activity. The polynucleotides may also be used for recombinant expression  
CC of the protein  
XX  
SQ Sequence 603 BP; 133 A; 179 C; 137 G; 154 T; 0 U; 0 Other;  
Query Match 100.0%; Score 19; DB 3; Length 603;  
Best Local Similarity 100.0%; Pred. No. 51;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CTTTACTTCATAGCTTTG 19  
Db 591 CTTTACTTCATAGCTTTG 573  
RESULT 7  
ABQ58527  
ID ABQ58527 standard; cDNA; 620 BP.  
XX  
AC ABQ58527;  
XX  
XX 02-AUG-2002 (first entry)  
XX  
XX Human colon cancer related nucleotide sequence SEQ ID NO:2222.  
XX  
DE Human; colon cancer; cancer; tissue profiling; forensic; mapping;  
KW genetic analysis; diagnostic; antisense therapy; gene; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200229086-A2.  
XX  
XX 11-APR-2002.  
XX  
XX 02-OCT-2001; 2001WO-US030732.  
XX  
XX 02-OCT-2000; 2000US-0237271P.  
XX  
XX (FARB ) BAYER CORP.  
XX  
XX Burgess C, Astle JH, Carroll E, Catino TJ, Dwivedi P, Molino GA;  
PI Thiagalingam A, Lewis ME;  
XX

XX WPI; 2002-426115/45.  
DR  
XX  
PT New isolated nucleic acid that is differentially expressed in cancer  
PT tissues useful for determining the presence of colon cancer in a cell or  
PT tissue type, and in antisense therapy.  
XX  
PS Claim 1; Fig 1; 796pp; English.  
XX  
CC ABQ56306 to ABQ60787 represent isolated nucleic acids (I) differentially  
CC expressed in cancer tissues. ABQ78993 to ABQ79004 represent proteins  
CC encoded by the ABQ60776 to ABQ60787 nucleic acid sequences. (I) can be  
CC used in antisense therapy. An antibody immunoreactive with a polypeptide  
CC encoded by (I) is useful for detecting cancer in a patient sample, and  
CC for detecting the presence or absence of a polynucleotide encoded by a  
CC nucleic acid which hybridises to (I) in a cell. A probe/primer derived  
CC from (I) can be used for determining the presence of a nucleic acid which  
CC hybridises to (I), and for determining the phenotype of cells in a sample  
CC of cells from a patient. (I) is useful for determining the presence of  
CC colon cancer in a cell or tissue type, for determining the presence or  
CC state of other type of cancer, in antisense therapy, to generate  
CC macroarrays on a solid surface, to identify a chromosome on which the  
CC corresponding gene resides, and in tissue profiling, forensics, genetic  
CC analysis, mapping and diagnostic applications. (I) can be used to raise  
CC antibodies, and to screen for peptide analogues and antagonists  
XX  
SQ Sequence 620 BP; 152 A; 154 C; 161 G; 142 T; 0 U; 11 Other;  
Query Match 100.0%; Score 19; DB 6; Length 620;  
Best Local Similarity 100.0%; Pred. No. 51;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CTTTACTTCATAGCTTTG 19  
Db 113 CTTTACTTCATAGCTTTG 131  
RESULT 8  
ABQ59698  
ID ABQ59698 standard; cDNA; 634 BP.  
XX  
AC ABQ59698;  
XX  
XX 02-AUG-2002 (first entry)  
XX  
XX Human colon cancer related nucleotide sequence SEQ ID NO:3393.  
XX  
DE Human; colon cancer; cancer; tissue profiling; forensic; mapping;  
KW genetic analysis; diagnostic; antisense therapy; gene; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200229086-A2.  
XX  
XX 11-APR-2002.  
XX  
XX 02-OCT-2001; 2001WO-US030732.  
XX  
XX 02-OCT-2000; 2000US-0237271P.  
XX  
XX (FARB ) BAYER CORP.  
XX  
XX Burgess C, Astle JH, Carroll E, Catino TJ, Dwivedi P, Molino GA;  
PI Thiagalingam A, Lewis ME;  
XX  
XX WPI; 2002-426115/45.  
XX  
PT New isolated nucleic acid that is differentially expressed in cancer  
PT tissues useful for determining the presence of colon cancer in a cell or  
PT tissue type, and in antisense therapy.  
XX  
PS Claim 1; Fig 1; 796pp; English.  
XX

CC ABQ56306 to ABQ60787 represent isolated nucleic acids (I) differentially  
 CC expressed in cancer tissues. ABQ78993 to ABQ79004 represent proteins  
 CC encoded by the ABQ60776 to ABQ60787 nucleic acid sequences. (I) can be  
 CC used in antisense therapy. An antibody immunoreactive with a polypeptide  
 CC encoded by (I) is useful for detecting cancer in a patient sample, and  
 CC for detecting the presence or absence of a polynucleotide encoded by a  
 CC nucleic acid which hybridises to (I) in a cell. A probe/primer derived  
 CC from (I) can be used for determining the presence of a nucleic acid which  
 CC hybridises to (I), and for determining the phenotype of cells in a sample  
 CC of cells from a patient. (I) is useful for determining the presence of  
 CC colon cancer in a cell or tissue type, for determining the presence or  
 CC state of other type of cancer, in antisense therapy, to generate  
 CC macroarrays on a solid surface, to identify a chromosome on which the  
 CC corresponding gene resides, and in tissue profiling, forensics, genetic  
 CC analysis, mapping and diagnostic applications. (I) can be used to raise  
 CC antibodies, and to screen for peptide analogues and antagonists

XX Sequence 634 BP; 153 A; 154 C; 168 G; 146 T; 0 U; 13 Other;

Query Match 100.0%; Score 19; DB 6; Length 634;  
 Best Local Similarity 100.0%; Pred. No. 51;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCTTTG 19  
 |||||  
 Db 117 CTTTACTTCATAGTCTTTG 135

RESULT 9  
 ID ABT22328 standard; DNA; 642 BP.  
 XX  
 AC ABT22328;  
 XX  
 DT 16-APR-2003 (first entry)  
 XX  
 DE Breast cancer marker gene SEQ ID No 701.

XX Cytostatic; vaccine; breast cancer marker gene; breast mass; immunogen;  
 KW chemotherapy; tumour burden; bait protein; two-hybrid; three-hybrid;  
 KW surrogate marker gene; pharmacodynamic marker gene; transgenic animal;  
 KW human; db.

XX Homo sapiens.

XX WO200285298-A2.

XX 31-OCT-2002.

XX 19-APR-2002; 2002WO-US012612.

XX 20-APR-2001; 2001US-0285163P.

XX (MILL-) MILLENNIUM PHARM INC.

XX Lillie J, Palermo A, Wang Y, Steinmann K, Elias J, Mertens M;

XX WPI; 2003-093053/08.

XX Novel isolated polypeptide encoded by breast cancer marker gene, useful  
 PT for diagnosing, staging, monitoring, prognosing and treating diseases  
 PT associated with breast cancer.

XX Disclosure; Page 187-188; 725pp; English.

XX The invention relates to an isolated polypeptide encoded by a breast  
 CC cancer marker gene comprising any of 1417 21-805 nucleotide sequences,  
 CC given in the specification. The methods of the invention are useful for  
 CC diagnosing patients having an identified breast mass or symptoms  
 CC associated with breast cancer, to diagnose breast cancer or its  
 CC precursors, and for monitoring the efficacy of treatment of a breast  
 CC cancer patient (e.g. efficacy of chemotherapy). The methods are also  
 CC useful for evaluating a patient before, after or during therapy, to

CC evaluate the reduction in a tumour burden. The breast cancer marker gene  
 CC proteins are useful as immunogens for raising antibodies, by immunising a  
 CC mammal with a breast cancer marker protein. The marker proteins are  
 CC useful as bait proteins in a two-hybrid or three-hybrid assay, to  
 CC identify other proteins which bind to or interact with the marker  
 CC proteins. The breast cancer marker genes are useful as surrogate marker  
 CC genes for one or more disorders, disease states or conditions leading to  
 CC disease states, in particular, breast cancers. The breast cancer marker  
 CC genes are useful as pharmacodynamic marker genes. An antibody which  
 CC selectively binds to a protein of a breast cancer marker gene is useful  
 CC for treating cancers, particularly breast cancers. The host cell of the  
 CC invention is useful for producing non-human transgenic animals. This  
 CC polynucleotide sequence represents one of the breast cancer marker genes  
 CC of the invention

XX Sequence 642 BP; 148 A; 160 C; 180 G; 142 T; 0 U; 12 Other;

Query Match 100.0%; Score 19; DB 10; Length 642;  
 Best Local Similarity 100.0%; Pred. No. 51;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCTTTG 19  
 |||||  
 Db 143 CTTTACTTCATAGTCTTTG 161

RESULT 10  
 ADO30035/c  
 ID ADO30035 standard; cDNA; 1074 BP.  
 XX  
 AC ADO30035;

XX 29-JUL-2004 (first entry)

XX Human GPCR RAI3 polynucleotide, SEQ ID NO:1137.

XX G protein-coupled receptor; GPCR; drug screening; diagnosis;  
 KW transgenic mouse; neurological disorder; adrenal gland disorder;  
 KW colon disorder; intestinal disorder; cardiovascular disorder;  
 KW muscular disorder; blood disorder; immune disorder; bone disorder;  
 KW joint disorder; metabolic disorder; nutritive disorder; cancer;  
 KW kidney disorder; liver disorder; lung disorder; breast disorder;  
 KW ovary disorder; uterus disorder; prostate disorder; testis disorder;  
 KW skin disorder; stomach disorder; pancreas disorder; spleen disorder;  
 KW thymus disorder; thyroid disorder; antiparkinsonian; antimanic;  
 KW cyostatic; antiinflammatory; vasotropic; antidiarrhoeic; antiarrhythmic;  
 KW CNS; central nervous system; respiratory; antidiarrhoeic; antidiabetic;  
 KW virucide; hepatotropic; antibacterial; antianaemic; antiseborrhoeic;  
 KW dermatological; antiulcer; antithyroid; antiallergic; anorectic;  
 KW immunosuppressive; nephrotropic; Gene therapy; GPCR modulator; human;  
 KW gene; ss.

XX Homo sapiens.

XX WO2004040000-A2.

XX 13-MAY-2004.

XX 09-SEP-2003; 2003WO-US028226.

XX 09-SEP-2002; 2002US-0409303P.

XX 09-APR-2003; 2003US-0461329P.

XX (PRIM-) PRIMAL INC.

XX Gaitanaris GA, Bergmann JE, Gragerov A, Hohmann J, Li F;

PI Madisen L, McIlwain KL, Pavlova MN, Vassilatis D, Zeng H;

XX WPI; 2004-390329/36.

XX P-PSDB; ADO29632.

XX Novel mammalian G protein coupled receptors, useful for identifying

PT compounds that modulates diagnosing and treating disease condition

associated with GPCR dysfunction e.g. autoimmune diseases, angina pectoris, Parkinson's disease.

Claim 151; SEQ ID NO 1137; 542pp; English.

The invention relates to human and mouse G protein-coupled receptors (GPCRs) and nucleic acids encoding them. The invention also relates to sequences at least 90% identical to the GPCR proteins and nucleic acids of the invention; methods of treating, preventing or diagnosing diseases associated with GPCRs of the invention; methods of screening for compounds useful in the treatment of GPCR-related diseases; a transgenic mouse comprising a GPCR gene of the invention; a mouse comprising a mutation in a GPCR transgene or in an endogenous GPCR gene; cells derived from the transgenic mice; kits comprising several mice, each of which has a mutation in a different GPCR gene of the invention; and kits comprising probes which hybridise to GPCR polynucleotides of the invention. The invention further discloses variants of the GPCR polypeptides and vectors comprising a GPCR nucleic acid. The GPCR nucleic acids and proteins may be used in the diagnosis, treatment or prevention of a wide variety of diseases including neurological disorders (e.g., Alzheimer's disease, depression, diabetic neuropathy, Parkinson's disease or schizophrenia); disorders of the adrenal gland; disorders of the colon or intestine (e.g., Crohn's disease, diarrhoea, food poisoning or irritable bowel syndrome); cardiovascular disorders (e.g., angina, cardiac arrhythmia or myocardial infarction); muscular disorders; blood disorders (e.g., anaemia or leukaemia); immune disorders (e.g., autoimmune disorders or AIDS); bone and joint disorders (e.g., osteoarthritis, rheumatoid arthritis, gout or osteoporosis); metabolic or nutritive disorders (e.g., obesity, enzyme deficiency-related diseases or vitamin deficiency-related diseases); and disorders of the kidney, liver, lung, breast, ovary, uterus, prostate, testis, skin, stomach, pancreas, spleen, thymus and thyroid (e.g., cancers). The present sequence represents a GPCR-encoding nucleic acid of the invention. Note: The full sequence data for this patent did not form part of the printed specification; those sequences not shown were obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences).

Sequence 1074 BP; 215 A; 316 C; 259 G; 284 T; 0 U; 0 Other;

Query Match 100.0%; Score 19; DB 12; Length 1074;  
Best Local Similarity 100.0%; Pred. No. 53;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGCTTTG 19  
|||||  
Db 1059 CTTTACTTCATAGCTTTG 1041

RESULT 11  
AAZ61776/c  
ID AAZ61776 standard; cDNA; 1212 BP.

XX AAC99709;

XX 27-MAR-2000 (first entry)

XX cDNA encoding human skin cell transmembrane protein, SEQ ID NO:249.

XX Skin; dermal papilla; keratinocyte; neonatal foreskin fibroblast;  
XX embryonic skin cell; keratinocyte stem cell; transit amplifying cell;  
XX secreted; transmembrane; inflammation; cancer; neurological disease;  
XX angiogenesis; tumour vascularisation; growth disorder;  
XX developmental disorder; skin wound; hair follicle disorder;  
XX anti-inflammatory; cytostatic; neuroprotective; vulnery; ss.

OS Homo sapiens.

XX WO9955865-A1.

XX 04-NOV-1999.

XX 29-APR-1999; 99WO-NZ000051.

XX

PR 29-APR-1998; 98US-00069726.  
PR 09-NOV-1998; 98US-00188930.  
XX (GENE-) GENESIS RES & DEV CORP LTD.

XX Strachan L, Sleeman M, Watson JD, Onrust R, Kumble A, Murison JG;  
PI WPI; 2000-072177/06.

XX Novel polynucleotides useful for the treatment of various conditions including wounds and cancer.

XX Claim 1; Page 155; 235pp; English.

XX The invention relates to novel nucleic acid sequences derived from rat dermal papilla, human keratinocytes and neonatal foreskin fibroblasts, and mouse embryonic skin, keratinocyte stem cells and transit amplifying cells. Polypeptides of the invention may be used to treat inflammation, cancer and neurological diseases. The proteins may be used to stimulate the growth and motility of keratinocytes, to inhibit the growth of cancer cells, to modulate angiogenesis and tumour vascularisation, to modulate skin inflammation, to modulate epithelial cell growth and to inhibit binding of HIV-1 to leukocytes. The invention may also be used to treat growth and developmental defects, skin wounds and hair follicle disorders. Sequences AAZ61606-261832 represent cDNA sequences derived from several mouse, rat or human skin cell types. Sequences AAZ61606-261649, AAZ61725-261765, AAZ61802-261811 and AAZ61826 encode proteins with an N-terminal signal sequence, indicating that the proteins are secreted. Sequences AAZ61650-261668, AAZ61766-261780, AAZ61812-261817 and AAZ61827-261829 encode proteins with one or more putative transmembrane domains

XX Sequence 1212 BP; 242 A; 350 C; 312 G; 308 T; 0 U; 0 Other;

Query Match 100.0%; Score 19; DB 3; Length 1212;  
Best Local Similarity 100.0%; Pred. No. 53;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGCTTTG 19  
|||||  
Db 1031 CTTTACTTCATAGCTTTG 1013

RESULT 12  
AAC99709/c  
ID AAC99709 standard; cDNA; 1212 BP.

XX AAC99709;

XX 08-MAR-2001 (first entry)

XX Skin cell cDNA, SEQ ID NO: 249.

XX Human; skin cell; cytostatic; antiinflammatory; anti-HIV; neurotropic;  
XX neuroprotective; vulnery; immunomodulatory; vaccine;  
XX keratinocyte growth stimulation; cancer; angiogenesis inhibition;  
XX inflammation; neurological disease; ss.

XX Homo sapiens.

XX WO200069884-A2.

XX 23-NOV-2000.

XX 15-MAY-2000; 2000WO-NZ000075.

XX 14-MAY-1999; 99US-00312283.

XX (GENE-) GENESIS RES & DEV CORP LTD.

XX Watson JD, Strachan L, Onrust R, Sleeman M, Kumble KD;  
PI Murison JG;

DR WPI; 2001-007495/01.  
DR P-PSDB; AAB55958.  
XX  
PT New isolated polynucleotide used in the identification of genetic  
PT disorders and encoding polypeptides used for treating inflammatory  
PT disease, cancer and neurological diseases.  
XX  
PS Claim 1; Page 210-211; 352pp; English.  
XX  
SS The present polynucleotide encodes a polypeptide which is expressed in  
CC mammalian skin cells. The polypeptide is useful for stimulating  
CC keratinocyte growth and motility, inhibiting the growth of cancer cells,  
CC modulating angiogenesis, inhibiting angiogenesis and vascularisation of  
CC tumours, modulating skin inflammation, stimulating the growth of  
CC epithelial cells, inhibiting the binding of human immunodeficiency virus  
CC (HIV)-1 to leukocytes, and treating inflammatory disease, cancer and  
CC neurological diseases. The polynucleotide can be used as a marker, in the  
CC identification of genetic disorders, and for the design of  
CC oligonucleotides for examining expression patterns  
XX  
SQ Sequence 1212 BP; 242 A; 350 C; 312 G; 308 T; 0 U; 0 Other;  
Query Match 100.0%; Score 19; DB 4; Length 1212;  
Best Local Similarity 100.0%; Pred. No. 53;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CTTTACTTCATAGTCCTTTG 19  
Db 1031 CTTTACTTCATAGTCCTTTG 1013  
RESULT 13  
ABL34861/c  
ID ABL34861 standard; cDNA; 1212 BP.  
XX  
AC ABL34861;  
XX  
DT 04-APR-2002 (first entry)  
XX  
DE Human cDNA isolated from skin cells SEQ ID NO: 249.  
XX  
KW Human; rat; mouse; skin cell; skin wound; cancer; growth defect;  
KW developmental defect; inflammatory disease; dermatological; vulneryary;  
KW immunomodulator; anti-inflammatory; cytostatic; neuroprotective; gene;  
KW ss.  
XX  
OS Homo sapiens.  
XX  
FN WO200190357-A1.  
XX  
PD 29-NOV-2001.  
XX  
XX 24-MAY-2001; 2001WO-NZ000099.  
XX  
XX 24-MAY-2000; 2000US-0206650P.  
XX  
XX 25-JUL-2000; 2000US-0221232P.  
XX  
PA (GENE-) GENESIS RES & DEV CORP LTD.  
XX  
XX Watson JD, Strachan L, Sleeman M, Onrust R, Murison JG;  
PI Kumble KD;  
XX  
DR WPI; 2002-122020/16.  
XX  
PT New polynucleotides and polypeptides encoded by the polynucleotides  
PT isolated from skin cells, useful for treating skin wounds, cancers,  
PT growth and developmental defects, inflammatory diseases, or for  
PT modulating immune responses.  
XX  
PS Claim 1; Page 175; 466pp; English.  
XX  
SS The present invention provides the protein and coding sequences of cDNAs  
CC isolated from human, murine and rat skin cell libraries. The sequences

CC can be used in the development of therapeutic agents useful in the  
CC treatment of skin diseases, including skin wounds, cancer, growth  
CC defects, developmental defects and inflammatory diseases. The proteins  
CC have important roles in the induction of hair growth, cell proliferation  
CC and cell-cell interaction, in maintaining tissue integrity, in wound  
CC healing and in modulating immune responses. The present sequence is a  
CC cDNA of the invention  
XX  
SQ Sequence 1212 BP; 242 A; 350 C; 312 G; 308 T; 0 U; 0 Other;  
Query Match 100.0%; Score 19; DB 6; Length 1212;  
Best Local Similarity 100.0%; Pred. No. 53;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CTTTACTTCATAGTCCTTTG 19  
Db 1031 CTTTACTTCATAGTCCTTTG 1013  
RESULT 14  
ACA04775/c  
ID ACA04775 standard; cDNA; 1228 BP.  
XX  
AC ACA04775;  
XX  
DT 28-MAY-2003 (first entry)  
XX  
DE cDNA encoding human membrane associated protein fragment #223.  
XX  
KW Human; ss; gene; microarray; membrane-associated protein; neuropathology;  
KW immunopathology; pancreatic disease; cancer; diabetes; hyperlipidaemia;  
KW pancreatic cholera; Alzheimer's disease; Huntington's disease; sarcoma;  
KW fibrocystic disease; leukaemia; adenocarcinoma; AIDS; allergy; anaemia;  
KW asthma; gout; dementia.  
XX  
OS Homo sapiens.  
XX  
PN US6492505-B1.  
XX  
PD 10-DEC-2002.  
XX  
XX 31-JAN-2000; 2000US-00495050.  
XX  
XX 01-FEB-1999; 99US-0118318P.  
XX  
PA (INCY-) INCYTE GENOMICS INC.  
XX  
XX Reddy R, Guegler KJ, Au-Young J;  
PI WPI; 2003-327324/31.  
XX  
DR Combination for research/diagnostic applications and for monitoring  
PT treatment of e.g., cancer, comprises polynucleotides comprising a  
PT fragment of gene encoding membrane-associated proteins, receptors or ion  
PT channels.  
XX  
PS Claim 1; Col 215-216; 147pp; English.  
XX  
CC The invention relates to a combination comprising several polynucleotide  
CC sequences comprising a fragment of gene encoding membrane-associated  
CC proteins, receptors or ion channels. The combination is useful as a  
CC probe, for research and diagnostic applications, for monitoring the  
CC expression of several expressed polynucleotides, in the diagnosis and  
CC monitoring of treatment of pancreatic disease, cancer, immunopathology or  
CC neuropathology, for investigating an individual's predisposition to the  
CC above disease, in genetic or gene expression analysis of polynucleotide  
CC sequences, to investigate cellular responses to infection or drug  
CC treatment, as hybridisable array elements in a microarray, to purify a  
CC subpopulation of mRNAs, cDNAs or genomic fragments in a sample, in  
CC diagnostics, prognostics and treatment regimens, in drug discovery and  
CC development, in toxicological and carcinogenicity studies, and in  
CC forensics or pharmacogenomics, to monitor the progression of disease, to  
CC monitor the efficacy of treatment, to diagnose the conditions of the

CC pancreas e.g. diabetes, pancreatic cholera, hyperlipidaemia or  
 CC fibrocystic disease, to diagnose a cancer e.g. leukaemia, adenocarcinoma  
 CC or sarcoma, to diagnose immunopathologies e.g. AIDS, allergies, anaemia,  
 CC asthma or gout, to diagnose neuropathologies e.g. Alzheimer's disease, of  
 CC dementia or Huntington's disease, to rapidly screen large numbers of  
 CC candidate drug molecules and as query sequences against GenBank,  
 CC SwissProt, BLOCKS and PRINTS databases. The combination is employed to  
 CC fine tune the treatment regimen and thus the expression patterns  
 CC associated with undesirable side effects are avoided. The present  
 CC sequence represents a cDNA encoding a fragment of gene encoding human  
 CC membrane-associated proteins, receptors or ion channels  
 XX  
 SQ Sequence 1228 BP; 272 A; 341 C; 283 G; 332 T; 0 U; 0 Other;

Query Match 100.0%; Score 19; DB 8; Length 1228;

Best Local Similarity 100.0%; Pred. No. 53;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGCTTTTG 19

Db 582 CTTTACTTCATAGCTTTTG 564

RESULT 15

AEA00112/c

ID AEA00112 standard; cDNA; 1460 BP.

XX AC AEA00112;

XX DT 28-JUL-2005 (first entry)

XX DE Human TAT143 cDNA SEQ ID NO:64.

XX ss; gene; tumor-associated antigen; cytostatic; breast tumor;  
 KW endocrine disease; gynecology and obstetrics; neoplasm; colon tumor;  
 KW gastrointestinal disease; rectal tumor; endometroid carcinoma;  
 KW genitourinary disease; renal tumor; lung tumor; respiratory disease;  
 KW ovary tumor; skin tumor; liver tumor.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX CDS 128..1201

XX FT /\*tag= a

XX FT /product= "TAT143"

XX PN US2005106644-A1.

XX XX 19-MAY-2005.

XX XX 08-SEP-2004; 2004US-00936626.

XX XX 20-JUN-2001; 2001US-0299500P.

XX PR 29-JUN-2001; 2001US-0301880P.

XX PR 18-SEP-2001; 2001US-0323268P.

XX PR 19-JUN-2002; 2002US-00177488.

XX PR 26-MAR-2004; 2004US-0557116P.

XX PR 04-AUG-2004; 2004US-0598899P.

XX XX (GETH ) GENENTECH INC.

XX XX Cairns B, Chen R, Frantz G, Hillan KJ, Koeppen H, Phillips HS;

XX PI Polakis P, Spencer SD, Smith V, Williams PM, Wu TD, Zhang Z;

XX PI Sliwowski M;

XX XX WPI; 2005-384304/39.

XX DR P-PSDB; AEA00190.

XX PT Novel isolated antibody capable of binding to tumor-associated antigenic  
 PT target polypeptide, useful for treating cell proliferative disorder e.g.  
 PT cancer.

XX XX Claim 1; SEQ ID NO 64; 337pp; English.

XX The invention relates to a novel isolated antibody binding to a  
 CC polypeptide having at least 80% sequence identity to a polypeptide having  
 CC any one of 76 fully defined 182-910 amino acid tumor-associated antigenic  
 CC target polypeptide (TAT) sequences (AEA00127-AEA00202) given in the  
 CC specification, a polypeptide having any one of (AEA00127-AEA00202),  
 CC lacking its associated signal peptide, or an extracellular domain of a  
 CC polypeptide having any one of (AEA00127-AEA00202). The polypeptide is  
 CC encoded by the nucleotide sequence having any one of (AEA00049-AEA00126).  
 CC An antibody of the invention has cytostatic activity. The antibody is  
 CC useful for inhibiting growth of a cell expressing TAT188, which involves  
 CC contacting the cell with the antibody. The cell is a cancer cell chosen  
 CC from breast, colon, rectum, endometrium, kidney, lung, ovary, skin and  
 CC liver cell. The cancer cell is a mammalian cell, preferably a TAT188  
 CC cell. The antibody is also useful for detecting the level of TAT188  
 CC polypeptide expressed in a test cell relative to a control cell, and for  
 CC detecting the level of TAT188 polypeptide or a polypeptide having at  
 CC least 80 % sequence identity to the TAT188 polypeptide sequence in a test  
 CC cell relative to a control cell. The antibody is useful for treating a  
 CC preventing a cell proliferative disorder associated with increased  
 CC expression or activity of a polypeptide having at least 80 % identity to  
 CC a TAT188 polypeptide sequence. The cell proliferative disorder is cancer.  
 CC The method of the invention is useful for inhibiting the growth of a  
 CC cancer cell. The present sequence encodes a polypeptide of the invention.

XX SQ Sequence 1460 BP; 302 A; 418 C; 362 G; 378 T; 0 U; 0 Other;

Query Match 100.0%; Score 19; DB 14; Length 1460;

Best Local Similarity 100.0%; Pred. No. 54;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGCTTTTG 19

Db 1186 CTTTACTTCATAGCTTTTG 1168

RESULT 16

AEA00632/c

ID AEA00632 standard; cDNA; 1460 BP.

XX AC AEA00632;

XX DT 28-JUL-2005 (first entry)

XX DE Human TAT143 cDNA sequence SeqID64.

XX antibody identification; tumor-associated antigen; cytostatic;  
 KW RNA interference; gene therapy; cell death; cancer; breast tumor;  
 KW colon tumor; rectal tumor; renal tumor; lung tumor; ovary tumor;  
 KW skin tumor; liver tumor; gene; ss; TAT143.

XX OS Homo sapiens.

XX PN US2005107595-A1.

XX PD 19-MAY-2005.

XX PF 10-SEP-2004; 2004US-00938061.

XX XX 20-JUN-2001; 2001US-0299500P.

XX PR 29-JUN-2001; 2001US-0301880P.

XX PR 18-SEP-2001; 2001US-0323268P.

XX PR 19-JUN-2002; 2002US-00177488.

XX PR 26-MAR-2004; 2004US-0557116P.

XX PR 04-AUG-2004; 2004US-0598899P.

XX XX (GETH ) GENENTECH INC.

XX XX Cairns B, Chen R, Frantz G, Hillan KJ, Koeppen H, Phillips HS;

XX PI Polakis P, Spencer SD, Smith V, Williams PM, Wu TD, Zhang Z;

XX PI Sakanaka C, Chuntharapai A, Reed CJ;

XX XX WPI; 2005-371577/38.

XX DR Claim 1; SEQ ID NO 64; 337pp; English.

P-PSDB; AEA00710.

Novel isolated antibody e.g. anti-E16 or anti-TAT112 antibody that binds to tumor-associated antigenic target polypeptide, useful for diagnosing or treating cancer.

Claim 1; SEQ ID NO 64; 96pp; English.

This invention relates to a novel isolated antibody, for example anti-E16 or anti-TAT112 antibody, that binds to a tumor-associated antigenic target polypeptide (TAT) and that lacks an associated signal peptide sequence. The invention may be useful for the development of compounds with a cytostatic activity acting as antagonists of the TAT188 polypeptide or RNA interference whilst the disclosed sequences may be useful for gene therapy. The invention is useful for inducing the death of a cell (such as a cancer cell chosen from breast, colon, rectum, endometrium, kidney, lung, ovary, skin and liver) to which it binds, inhibiting proliferation or promoting cell death of a cell expressing TAT188. In addition, the invention may be useful for detecting the level of TAT188 polypeptide in a test cell relative to a control cell, or treating or preventing a cell proliferative disorder associated with increased expression of TAT188. The novel antibody of the invention is useful for inhibiting the growth of a cancer cell and may be useful for diagnosing or treating cancer. The present sequence is that of the human TAT143 cDNA which encodes a protein against which an antibody of the invention may be targeted.

Sequence 1460 BP; 302 A; 418 C; 362 G; 378 T; 0 U; 0 Other;

Query Match 100.0%; Score 19; DB 14; Length 1460;  
Best Local Similarity 100.0%; Pred. No. 54;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCCTTTG 19  
|||||  
Db 1186 CTTTACTTCATAGTCCTTTG 1168

RESULT 17  
AAF58615/c  
ID AAF58615 standard; cDNA; 1619 BP.  
XX AAF58615;  
XX 24-APR-2001 (first entry)  
XX Human RECAP polynucleotide, SEQ ID NO: 43.  
XX Human, RECAP; receptors and associated proteins; cerebroprotective; neurotropic; neuroprotective; anticonvulsant; antiparkinsonian; anti-HIV; antidiabetic; immunostimulant; immunomodulator; antiinflammatory; antichyroid; immunosuppressive; nephrotropic; antigout; thyromimetic; cytostatic; antibacterial; virucide; fungicide; protozoacide; antiarteriosclerotic; hepatotropic; gene therapy; infection; cancer; ss.  
XX Homo sapiens.  
XX WO200107612-A2.  
XX 01-FEB-2001.  
XX 21-JUL-2000; 2000WO-US020035.  
XX 21-JUL-1999; 99US-0145232P.  
XX 07-OCT-1999; 99US-0158578P.  
XX 12-NOV-1999; 99US-0165192P.  
XX (INCY-) INCYTE GENOMICS INC.  
XX Au-Young J, Bandman O, Tang YT, Yue H, Azimzai Y, Burford N;  
XX Baughn MR, Lu DAM, Hillman JL, Patterson C, Lal P;  
XX WPI; 2001-169554/17.

P-PSDB; AAB68891.

Novel receptors and associated proteins for diagnosis and treatment of PT neurological disorders, immunological disorders including autoimmune/ PT inflammatory disorders and cell proliferative disorders such as cancer.

XX Example 5; Page 126-127; 128pp; English.

The present sequence encodes a human RECAP (receptors and associated proteins) polypeptide. RECAP polynucleotides and polypeptides are useful in the diagnosis, treatment and prevention of neurological disorders such as stroke, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease, Down's syndrome, amyotrophic lateral sclerosis, multiple sclerosis, bacterial and viral meningitis, CJD (Creutzfeldt-Jakob disease), GSS (Gerstmann -Straussler-Scheinker syndrome); immunological disorders, including autoimmune/inflammatory disorders such as AIDS, DiGeorge's syndrome, severe combined immunodeficiency disease (SCID), Chediak-Higashi syndrome, Cushing's disease, Addison's disease, autoimmune thyroiditis, Crohn's disease, diabetes mellitus, Good pasture's syndrome, gout, Grave's disease, Hashimoto's thyroiditis, Sjogren's syndrome, Werner's syndrome, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections; and cell proliferation disorders such as arteriosclerosis, atherosclerosis, cirrhosis, hepatitis and cancer

Sequence 1619 BP; 331 A; 463 C; 400 G; 425 T; 0 U; 0 Other;

Query Match 100.0%; Score 19; DB 4; Length 1619;  
Best Local Similarity 100.0%; Pred. No. 54;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCCTTTG 19  
|||||  
Db 1178 CTTTACTTCATAGTCCTTTG 1160

RESULT 18  
AAZ90056/c  
ID AAZ90056 standard; cDNA; 1718 BP.  
XX AAZ90056;  
XX 09-MAY-2000 (first entry)  
XX Hydrophobic domain containing protein clone HP10549 nucleotide sequence.  
XX Hydrophobic domain; clone HP10549; nutritional supplement; SCID; HIV; cell proliferation; immune stimulant; immune deficiency; tumour; pain; rheumatoid arthritis; insulin dependent diabetes mellitus; fertility; myasthenia gravis; haematopoiesis regulator; tissue growth; depression; anti-inflammatory; infection; bodily characteristic; ss.  
XX Homo sapiens.  
XX WO200000506-A2.  
XX 06-JAN-2000.  
XX 18-JUN-1999; 99WO-JP003242.  
XX 26-JUN-1998; 98JP-00180008.  
XX (SAGA) SAGAMI CHEM RES CENT.  
XX (PROT-) PROTEGENE INC.  
XX Kato S, Kimura T;  
XX WPI; 2000-160665/14.  
XX P-PSDB; AAY78809.  
XX Novel human proteins having hydrophobic domains used for research and diagnostic purposes.



PS Claim 4; Page 111-113; 117pp; English.

XX This sequence represents the hydrophobic domain containing protein, clone  
CC HP10549 nucleotide sequence. The sequence is isolated from a human  
CC stomach cancer cell line. The invention relates to human proteins with  
CC hydrophobic domains, the DNA and the cDNA encoding them. The  
CC polynucleotides and proteins are predicted to have biological activities  
CC which make them suitable for treating, preventing or ameliorating medical  
CC conditions in humans and animals. Suggested activities include  
CC nutritional activity (nutritional source or supplement); cytokine and  
CC cell proliferation/differentiation activity; immune stimulating (e.g. as  
CC vaccines) or suppressing activity (e.g. to treat various immune  
CC deficiencies such as SCIDS or HIV, connective tissue disease, systemic  
CC lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary  
CC inflammation, Guillain-Barre syndrome, myasthenia gravis, graft-versus-host disease  
CC and autoimmune inflammatory eye disease, as well as asthma, allergies and  
CC organ transplantation); haematopoiesis regulating activity (e.g. in  
CC treatment of myeloid or lymphoid cell deficiencies); tissue growth  
CC activity (e.g. wound healing and tissue repair, ulcers, burns,  
CC periodontal disease); activin/inhibin activity; chemotactic/chemokinetic  
CC activity; haemostatic and thrombolytic activity (e.g. treating  
CC haemophilias); receptor/ligand activity; anti-inflammatory activity; and  
CC tumour inhibition activity. The polynucleotides are also stated to be  
CC useful for gene therapy. Other activities include inhibiting infections  
CC caused by bacteria, fungi, viruses and other parasites (e.g. Hepatitis,  
CC malaria); effecting bodily characteristics such as, e.g. weight, colour,  
CC skin, effecting biorhythms or circadian cycles; enhancing fertility;  
CC treatment of depression; treatment of pain; hormonal or endocrine  
CC activity. The polynucleotides may also be used for recombinant expression  
CC of the protein

XX SQ Sequence 1718 BP; 381 A; 495 C; 377 G; 465 T; 0 U; 0 Other;

Query Match 100.0%; Score 19; DB 3; Length 1718;

Best Local Similarity 100.0%; Pred. No. 54;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19

Db 602 CTTTACTTCATAGTCTTTG 584

RESULT 19

ABT10173/C

ID ABT10173 standard; cDNA; 2302 BP.

XX AC ABT10173;

XX 04-DEC-2002 (first entry)

XX Human breast cancer associated coding sequence SEQ ID NO: 307.

XX Human; breast specific gene; breast cancer; differential expression;

XX cytostatic; gene therapy; gene; ss.

XX Homo sapiens.

XX WO200259271-A2.

XX 01-AUG-2002.

XX 25-JAN-2002; 2002WO-US002176.

XX 25-JAN-2001; 2001US-0263757P.

XX 25-APR-2001; 2001US-0286090P.

XX 23-MAY-2001; 2001US-0292517P.

XX (GENE-) GENE LOGIC INC.

XX Orr MS, Nation M, Diggans JC, Zeng W;

XX WPI; 2002-674803/72.

XX

PT Diagnosing breast cancer in a patient comprises detecting the level of  
PT gene expression in cell or tissue samples, where a differential gene  
PT expression is indicative of breast cancer.

XX PS Claim 1; SEQ ID NO 307; 260pp + Sequence Listing; English.

XX The present invention relates to methods of diagnosing breast cancer in a  
CC patient, which comprise detecting the level of expression in a tissue  
CC sample of two or more genes selected from those shown in ABT09867-  
CC ABT1112, where a differential expression of the genes indicates breast  
CC progression, and in monitoring treatment of breast cancer in patients.  
CC The methods are also useful as a screening tool for agents that modulate  
CC the onset or progression of breast cancer. The breast cancer genes may be  
CC used as diagnostic markers for the prediction or identification of the  
CC malignant state of breast tissue, for confirming the type and progression  
CC of cancer, and for drug screening and assays. The present sequence is a  
CC coding sequence of the invention. Note: The sequence data for this patent  
CC did not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub.published\_pct\_sequences

XX SQ Sequence 2302 BP; 494 A; 666 C; 533 G; 609 T; 0 U; 0 Other;

Query Match 100.0%; Score 19; DB 6; Length 2302;

Best Local Similarity 100.0%; Pred. No. 55;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19

Db 1158 CTTTACTTCATAGTCTTTG 1140

RESULT 20

ACC58386/C

ID ACC58386 standard; cDNA; 2302 BP.

XX AC ACC58386;

XX 26-AUG-2003 (first entry)

XX Human GPCR-like retinoic acid-induced gene 1.

XX Human; retinoic acid-induced gene 1; RAIG1; feeding; fasting; GPCR;  
XX receptor; G-protein coupled receptor; anorectic; antidiabetic;  
XX antidepressant; immunomodulator; transgenic; gene therapy; gene; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 100..1173

XX /\*tag= a

XX /product= "GPCR-like RAIG1"

XX WO2003016553-A2.

XX 27-FEB-2003.

XX 20-AUG-2002; 2002WO-US026510.

XX 20-AUG-2001; 2001US-0313940P.

XX (GETH ) GENENTECH INC.

XX (CURA-) CURAGEN CORP.

XX Lewin DA, Stewart TA;

XX WPI; 2003-278580/27.

XX P-PSDB; ABR42649.

XX New G-protein coupled receptor-like retinoic acid induced gene 1 (GPCR-  
XX like RAIG1) polypeptide and gene, useful for diagnosing or treating

PT metabolic disorders, e.g. obesity, anorexia, cachexia or diabetes.  
XX Claim 10; Page 16-17; 150pp; English.  
PS  
XX  
CC The present sequence is that of human G-protein coupled receptor-like  
CC retinoic acid induced gene 1 (GPCR-like RAIG1). This is the human  
CC homologue of murine GPCR-like RAIG1. The murine gene was shown to be  
CC differentially regulated during fasting-feeding cycles in mice, with  
CC moderate induction early in fasting, down-regulation with extended  
CC fasting and 4-fold up-regulation with feeding in recovery from fasting.  
CC The differentially expressed gene, its mRNA, and the encoded protein, can  
CC each be manipulated to detect and treat metabolic disorders associated  
CC with up- or down-regulation of GPCR-like RAIG1 activity, such as obesity,  
CC anorexia, cachexia or diabetes  
XX  
SQ Sequence 2302 BP; 494 A; 666 C; 533 G; 609 T; 0 U; 0 Other;  
Query Match 100.0%; Score 19; DB 8; Length 2302;  
Best Local Similarity 100.0%; Pred. No. 55;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CTTTACTTCATAGTCTTTG 19  
Db 1158 CTTTACTTCATAGTCTTTG 1140  
RESULT 21  
ADD93240/c  
ID ADD93240 standard; cDNA; 2302 BP.  
XX  
AC ADD93240;  
DT  
XX 29-JAN-2004 (first entry)  
DE RAIG1 coding sequence.  
XX  
XX ss; gene; retinoic acid-inducible gene 1 ; RAIG1; orphan ;  
KW G-protein coupled receptor; GPCR; chromosome 12; GPCR5B; carcinoma ;  
KW vaccine; breast cancer; pancreatic cancer; lung cancer; liver cancer;  
KW ovarian cancer; colon cancer ; osteosarcoma.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 100..1173  
FT /\*tag= a  
FT /product= "RAIG1"  
XX  
XX WO2003087832-A2.  
XX  
XX 23-OCT-2003.  
XX  
XX 10-APR-2003; 2003WO-GB001587.  
XX  
XX 11-APR-2002; 2002GB-00008331.  
XX  
XX 17-SEP-2002; 2002GB-00021538.  
XX  
XX (OXFO-) OXFORD GLYSCSCIENCES UK LTD.  
XX  
XX Terrett JA;  
XX  
XX WPI; 2003-845382/78.  
XX  
XX P-PSDB; ADD93239.  
XX  
XX GENBANK; AF095448.  
XX  
XX Screening, diagnosing and/or treating carcinoma, including breast,  
XX pancreatic, lung, liver, ovarian and colon cancer by detecting the change  
XX in expression or activity of an RAIG1 polypeptide or encoding nucleic  
XX acid molecule.  
XX  
XX Claim 1; Fig 2; 43pp; English.  
XX  
XX This sequence encodes retinoic acid-inducible gene 1 (RAIG1) polypeptide.

CC RAIG1 is an orphan G-protein coupled receptor (GPCR) located on  
CC chromosome 12. RAIG1 shows a restricted expression pattern compared to a  
CC related receptor, GPCR5B, which is widely expressed in peripheral and  
CC central tissues. The RAIG1 cDNA and polypeptide sequence may be used in  
CC the method of the invention for screening for and/or diagnosis of  
CC carcinoma in a subject, and/or monitoring the effectiveness of carcinoma  
CC therapy. The method comprises detecting and/or quantifying in a  
CC biological sample obtained from the subject an RAIG1 polypeptide and a  
CC nucleic acid molecule. The RAIG1 polypeptide and nucleic acid molecule  
CC are useful in the manufacture of a medicament for the treatment of  
CC carcinoma, where the composition is a vaccine. An agent which interacts  
CC with or causes change in the expression or activity of an RAIG1  
CC polypeptide or nucleic acid molecule, is also useful in the manufacture  
CC of a medicament for the treatment of carcinoma that is breast cancer,  
CC pancreatic cancer, lung cancer, liver cancer, ovarian cancer, colon  
CC cancer and/or osteosarcoma. They can also be used in the diagnosis and  
CC screening of such carcinomas.  
XX  
SQ Sequence 2302 BP; 494 A; 666 C; 533 G; 609 T; 0 U; 0 Other;  
Query Match 100.0%; Score 19; DB 10; Length 2302;  
Best Local Similarity 100.0%; Pred. No. 55;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CTTTACTTCATAGTCTTTG 19  
Db 1158 CTTTACTTCATAGTCTTTG 1140  
RESULT 22  
ADL24773/c  
ID ADL24773 standard; DNA; 2302 BP.  
XX  
AC ADL24773;  
DT  
XX 20-MAY-2004 (first entry)  
XX  
DE Intestinal epithelium/peyer's patch M cell-associated DNA sequence #171.  
XX  
XX intestinal epithelium cell development; peyer's patch M cell development;  
XX inflammatory bowel disease; glutenenteropathy; infectious disease;  
KW autoimmune disease; haemolytic anaemia; rheumatoid arthritis; dermatitis;  
KW Grave's disease; multiple sclerosis; allergy; asthma; diabetic mellitus;  
KW immune system disorder; hypersensitivity; anaphylaxis;  
KW blood group incompatibility; ds; human.  
XX  
XX Homo sapiens.  
XX  
XX WO200280852-A2.  
XX  
XX 17-OCT-2002.  
XX  
XX 04-APR-2002; 2002WO-US010873.  
XX  
XX 04-APR-2001; 2001US-0281416P.  
XX  
XX (DIGI-) DIGITAL GENE TECHNOLOGIES INC.  
XX  
XX Brayden DJ, Byrne D, O'mahony DJ, Evans CF, Mah SP, Lo DD;  
XX  
XX WPI; 2003-075470/07.  
XX  
XX Novel isolated or purified polypeptide encoded by genes associated with  
XX intestinal epithelium or M cell development, differentiation or function,  
XX useful for treating autoimmune diseases and infectious diseases.  
XX  
XX Claim 1; SEQ ID NO 283; 152pp; English.  
XX  
XX The invention comprises DNA sequences which are associated with  
XX intestinal epithelium and peyer's patch M cells. The DNA sequences of the  
XX invention are useful for assessing, modifying, modulating or regulating  
XX intestinal epithelium or M cell development. The DNA sequences of the  
XX invention are also useful in the treatment of: inflammatory bowel

CC disease, gluteenteropathy, infectious diseases, autoimmune diseases  
CC (e.g. haemolytic anaemia, rheumatoid arthritis, dermatitis, Grave's  
CC disease, multiple sclerosis, allergy, asthma and diabetic mellitus),  
CC diseases or disorders of the immune system, hypersensitivity,  
CC anaphylaxis, and blood group incompatibility. The present nucleic acid  
CC represents an intestinal epithelium/peyer's patch M cell-associated DNA  
CC sequence of the invention. NOTE: The present sequence is not shown in the  
CC specification, but has been retrieved from the WIPO website.  
XX  
SQ Sequence 2302 BP; 494 A; 566 C; 533 G; 609 T; 0 U; 0 Other;

Query Match 100.0%; Score 19; DB 10; Length 2302;  
Best Local Similarity 100.0%; Pred. No. 55;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGCTTTG 19  
Db 1158 CTTTACTTCATAGCTTTG 1140

## RESULT 23

ABT31923/c

ID ABT31923 standard; DNA; 2316 BP.

XX AC ABT31923;

XX 01-MAY-2003 (first entry)

XX Human breast cancer / ovarian cancer related coding sequence #30.

XX Human; gene; ds; cytosstatic; breast cancer; ovarian cancer.

XX Homo sapiens.

XX WO200300012-A2.

XX 03-JAN-2003.

XX 21-JUN-2002; 2002WO-US019773.

XX 21-JUN-2001; 2001US-0300159P.

XX 27-JUN-2001; 2001US-0301351P.

XX (MILL-) MILLENNIUM PHARM INC.

XX Veiby OP;

XX WPI; 2003-267848/26.

XX P-PSDB; ABJ37054.

XX Determining the presence of breast cancer in an individual, involves  
PT using specific polynucleotide markers.

XX Disclosure; Page 163; 233pp; English.

XX The invention comprises a method for assessing whether a patient is  
XX afflicted with breast cancer or ovarian cancer. The method involves the  
XX use of specific DNA markers. The method of the invention is useful in the  
XX detection and treatment of ovarian and breast cancer. DNA sequences  
XX ABT31894 - ABT31949 encode human breast/ovarian cancer-related proteins  
XX

SQ Sequence 2316 BP; 502 A; 670 C; 535 G; 609 T; 0 U; 0 Other;

Query Match 100.0%; Score 19; DB 10; Length 2316;  
Best Local Similarity 100.0%; Pred. No. 55;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGCTTTG 19  
Db 1164 CTTTACTTCATAGCTTTG 1146

## RESULT 24

AAH14688/c  
ID AAH14688 standard; cDNA; 2446 BP.  
XX  
AC AAH14688;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
DE Human cDNA sequence SEQ ID NO:12388.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

XX EPI074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-00116126.

XX 29-JUL-1999; 99JP-00248036.

XX 27-AUG-1999; 99JP-00300253.

XX 11-JAN-2000; 2000JP-00118776.

XX 02-MAY-2000; 2000JP-00183767.

XX 09-JUN-2000; 2000JP-00241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

XX Claim 8; SEQ ID NO 12388; 2537pp + Sequence Listing; English.

XX The present invention describes primer sets for synthesising 5602 full-length cDNAs defined in the specification. Where a primer set comprises:  
XX (a) an oligo-dT primer and an oligonucleotide complementary to the  
XX complementary strand of a polynucleotide which comprises one of the 5602  
XX nucleotide sequences defined in the specification, where the  
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
XX of an oligonucleotide comprising a sequence complementary to the  
XX complementary strand of a polynucleotide which comprises a 5'-end  
XX sequence and an oligonucleotide comprising a sequence complementary to a  
XX polynucleotide which comprises a 3'-end sequence, where the  
XX oligonucleotide comprises at least 15 nucleotides and the combination of  
XX the 5'-end sequence/3'-end sequence is selected from those defined in the  
XX specification. The primer sets can be used in antisense therapy and in  
XX gene therapy. The primers are useful for synthesising polynucleotides,  
XX particularly full-length cDNAs. The primers are also useful for the  
XX detection and/or diagnosis of the abnormality of the proteins encoded by  
XX the full-length cDNAs. The primers allow obtaining of the full-length  
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
XX AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893  
XX represent human amino acid sequences; and AAH13629 to AAH13632 represent  
XX oligonucleotides, all of which are used in the exemplification of the  
XX present invention

SQ Sequence 2446 BP; 507 A; 723 C; 572 G; 644 T; 0 U; 0 Other;

Query Match 100.0%; Score 19; DB 4; Length 2446;  
Best Local Similarity 100.0%; Pred. No. 56;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGCTTTG 19  
Db 1312 CTTTACTTCATAGCTTTG 1294

RESULT 25  
AB242832/c  
ID AB242832 standard; DNA; 2456 BP.  
XX  
XX  
AC AB242832;  
XX  
XX  
DT 04-MAR-2003 (first entry)  
DE  
DE Human G protein-coupled receptor RAIG1 nucleotide SEQ ID NO:453.  
XX  
XX  
KW G protein-coupled receptor; GPCR; antigenic peptide; gene therapy;  
KW G protein-coupled receptor modulator; antibody; immune-related disease;  
KW growth-related disease; cell regeneration-related disease; AIDS; cancer;  
KW immunological-related cell proliferative disease; autoimmune disease;  
KW Alzheimer's disease; atherosclerosis; infection; osteoarthritis; allergy;  
KW osteoporosis; cardiomyopathy; inflammation; Crohn's disease; diabetes;  
KW graft versus host disease; Parkinson's disease; multiple sclerosis; pain;  
KW psoriasis; anxiety; depression; schizophrenia; dementia; memory loss;  
KW mental retardation; epilepsy; asthma; tuberculosis; obesity; nausea;  
KW hypertension; hypotension; renal disorder; rheumatoid arthritis; trauma;  
KW ulcer; gene; ds.  
XX  
OS Homo sapiens.  
XX  
XX WO200261087-A2.  
XX  
XX PD 08-AUG-2002.  
XX  
XX PF 19-DEC-2001; 2001WO-US050107.  
XX  
XX PR 19-DEC-2000; 2000US-0257144P.  
XX  
XX PA (LIFE-) LIFESPAN BIOSCIENCES INC.  
XX  
XX PI Burmer GC, Roush CL, Brown JP;  
XX  
XX DR WPI; 2003-046718/04.  
XX  
XX DR P-PSDB; ABP81984.  
XX  
XX PT New isolated antigenic peptides e.g., for G protein-coupled receptors  
XX (GPCR), useful for diagnosing and designing drugs for treating conditions  
XX in which GPCRs are involved, e.g. AIDS, Alzheimer's disease, cancer or  
XX autoimmune diseases.  
XX  
XX PS Disclosure; Fig 1; 523pp; English.  
XX  
XX CC The present invention describes antigenic peptides (I) comprising: (a)  
XX any one of 1601 sequences (see ABP82019 to ABP83619) of 12-24 amino  
XX acids. Also described: (1) an assay for the detection of a particular G  
XX protein-coupled receptor (GPCR) or a candidate polypeptide in a sample;  
XX and (2) an isolated antibody having high specificity and high affinity or  
XX avidity for a particular GPCR. (I) can be used as GPCR modulators and in  
XX gene therapy. The antigenic peptides for GPCRs are useful in detecting an  
XX antibody against a particular GPCR, and in the production of specific  
XX antibodies. The peptides and antibodies are also useful for detecting the  
XX presence or absence of corresponding GPCRs. The antigenic peptides for  
XX GPCRs and antibodies are useful for diagnosing and designing drugs for  
XX treating immune-related diseases, growth-related diseases, cell  
XX regeneration-related disease, immunological-related cell proliferative  
XX diseases, or autoimmune diseases, e.g. AIDS, Alzheimer's disease,  
XX atherosclerosis, bacterial, fungal, protozoan or viral infections,  
XX osteoarthritis, osteoporosis, cancer, cardiomyopathy, chronic and acute  
XX inflammation, allergies, Crohn's disease, diabetes, graft versus host  
XX disease, Parkinson's disease, multiple sclerosis, pain, psoriasis,  
XX anxiety, depression, schizophrenia, dementia, mental retardation, memory  
XX loss, epilepsy, asthma, tuberculosis, obesity, nausea, hypertension,  
XX hypotension, renal disorders, rheumatoid arthritis, trauma, ulcers, or  
XX any other disorder in which GPCRs are involved. The antibodies may be  
XX used in immunoassays and immunodiagnosis. AB242523 to AB242869 encode  
XX GPCR proteins given in ABP81675 to ABP82018, which are used in the  
XX exemplification of the present invention  
XX  
XX SQ Sequence 2456 BP; 519 A; 722 C; 571 G; 644 T; 0 U; 0 Other;

Query Match 100.0%; Score 19; DB 8; Length 2456;  
Best Local Similarity 100.0%; Pred. No. 56;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CTTTACTTCATAGTCTTTG 19  
DB 1312 CTTTACTTCATAGTCTTTG 1294  
RESULT 26  
ACC72695/c  
ID ACC72695 standard; cDNA; 2456 BP.  
XX  
XX AC ACC72695;  
XX  
XX DT 09-JUL-2003 (first entry)  
XX  
XX DE Human cancer related protein encoding cDNA SEQ ID NO:34.  
XX  
XX KW Human; cancer; diagnosis; screening; modulator; leukaemia; ischaemia;  
XX heart disease; atherosclerosis; endometriosis; gene; ss.  
XX  
XX OS Homo sapiens.  
XX  
XX PN WO2003025138-A2.  
XX  
XX PD 27-MAR-2003.  
XX  
XX PF 17-SEP-2002; 2002WO-US029560.  
XX  
XX PR 17-SEP-2001; 2001US-0323469P.  
XX PR 20-SEP-2001; 2001US-0323887P.  
XX PR 13-NOV-2001; 2001US-0350666P.  
XX PR 08-FEB-2002; 2002US-0355145P.  
XX PR 08-FEB-2002; 2002US-0355257P.  
XX PR 12-APR-2002; 2002US-0372246P.  
XX  
XX PA (EOSB-) EOS BIOTECHNOLOGY INC.  
XX  
XX PI Afar D, Aziz N, Gish KC, Hevezi PA, Mack DH, Wilson KE;  
XX Zlotnik A;  
XX  
XX DR WPI; 2003-354600/33.  
XX DR P-PSDB; ABR58568.  
XX  
XX PT New genes that are up-regulated or down-regulated in cancers, useful as  
XX markers for diagnosing e.g. cancer, ischemia or heart diseases, or as  
XX therapeutic targets for screening drugs for treating these diseases.  
XX  
XX PS Claim 8; Page 643-644; 767pp; English.  
XX  
XX CC The present invention describes an isolated nucleic acid molecule, which  
XX comprises the sequence of any of the genes that are up-regulated or down-  
XX regulated in specific cancers (e.g. about 1031 genes up-regulated in  
XX acute lymphocytic leukemia). ACC72641 to ACC72860 represent cancer  
XX related gene nucleotide sequences which encode the proteins given in  
XX ABR58521 to ABR58709. Also described: (1) determining the presence or  
XX absence of a pathological cell in a patient; (2) an expression vector  
XX comprising a nucleic acid molecule described above; (3) a host cell  
XX comprising the vector; (4) an isolated polypeptide, which is encoded by  
XX the nucleic acid; (5) an antibody that specifically binds the polypeptide  
XX of (4); (6) specifically targeting a compound to a pathological cell in a  
XX patient by administering to the patient the antibody above; and (7) a  
XX drug screening assay. The nucleic acid is useful as diagnostic markers or  
XX therapeutic targets. In particular, the nucleic acid is useful for  
XX diagnosing a pathology, e.g. cancer (e.g. cancer of the bone marrow,  
XX bladder, brain, breast, cervix, colon/rectum, kidney, lung, ovary,  
XX pancreas, prostate, skin and uterus), wounds, ischaemia, heart diseases,  
XX atherosclerosis and endometriosis. The nucleic acid is also useful in  
XX drug screening, particularly for identifying agents for treating these  
XX pathologies

SQ Sequence 2456 BP; 522 A; 720 C; 571 G; 643 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 19; DB 10; Length 2456;  
 Best Local Similarity 100.0%; Pred. No. 56;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 CTTTACTTCATAGTCTTTG 19  
 Db 1312 CTTTACTTCATAGTCTTTG 1294  
 RESULT 27  
 ADN39301/c  
 ID ADN39301 standard; cDNA; 2456 BP.  
 XX AC ADN39301;  
 XX DT 17-JUN-2004 (first entry)  
 XX DE Cancer/angiogenesis/fibrosis-related nucleic acid, SEQ ID NO:619.  
 XX KW Human; differential expression; cancer; angiogenic disorder;  
 KW fibrotic disorder; psoriasis; ischaemia; heart disease; atherosclerosis;  
 KW inflammatory disease; autoimmune disease;  
 KW retinal neovascularisation syndrome; scarring; uterine fibroid;  
 KW detection; diagnosis; prognosis; drug screening; drug targeting;  
 KW wound healing; contraception; cytostatic; cardiant; immunomodulatory;  
 KW vulnery; gene therapy; vaccine; gene; ss.  
 XX OS Homo sapiens.  
 XX PN W02003042661-A2.  
 XX PD 22-MAY-2003.  
 XX PF 13-NOV-2002; 2002WO-US036810.  
 XX PR 13-NOV-2001; 2001US-0350666P.  
 PR 21-NOV-2001; 2001US-0332464P.  
 PR 29-NOV-2001; 2001US-0334393P.  
 PR 03-DEC-2001; 2001US-0335394P.  
 PR 14-DEC-2001; 2001US-0340376P.  
 PR 08-JAN-2002; 2002US-0347211P.  
 PR 10-JAN-2002; 2002US-0347349P.  
 PR 08-FEB-2002; 2002US-035250P.  
 PR 13-FEB-2002; 2002US-0356714P.  
 PR 20-FEB-2002; 2002US-0359077P.  
 PR 29-MAR-2002; 2002US-036809P.  
 PR 04-APR-2002; 2002US-0370110P.  
 PR 12-APR-2002; 2002US-0372246P.  
 PR 05-JUN-2002; 2002US-0386614P.  
 PR 16-JUL-2002; 2002US-0396839P.  
 PR 22-JUL-2002; 2002US-0397775P.  
 PR 22-JUL-2002; 2002US-0397845P.  
 PR 09-SEP-2002; 2002US-0409450P.  
 XX PA (EOSB-) EOS BIOTECHNOLOGY INC.  
 XX PI Afar D, Aziz N, Ginsburg WM, Gish KC, Glynn R, Hevezi PA;  
 PI Mack DH, Murray R, Watson SR, Wilson KE, Zlotnik A;  
 XX WIPI: 2003-468649/44.  
 DR P-PSDB; ADN39302.  
 XX Determining the presence or absence of a pathological cell in a patient,  
 PT useful for diagnosing, prognosing or treating cancer, comprises detecting  
 PT a nucleic acid in a biological sample.  
 XX Claim 8; SEQ ID NO 619; 1385pp; English.  
 CC The invention relates to nucleic acids and proteins (ADN38683-ADN40064)  
 CC whose expression is upregulated or downregulated in specific cancers or  
 CC other diseases such as angiogenic or fibrotic disorders, and to methods

CC of determining the presence or absence of a pathological cell in a  
 CC patient by detecting a nucleic acid at least 80% identical to those of  
 CC the invention or by detecting a polypeptide of the invention. The  
 CC invention also relates to expression vectors and host cells comprising a  
 CC nucleic acid of the invention; antibodies which specifically bind a  
 CC polypeptide of the invention; use of such antibodies for drug targeting;  
 CC and methods of screening for modulators of activity or expression of the  
 CC polypeptides and nucleic acids. The nucleic acids, polypeptides,  
 CC antibodies and methods are useful for diagnosing, prognosing and treating  
 CC cancer and other conditions such as psoriasis, ischaemia, heart disease,  
 CC atherosclerosis, inflammatory diseases, autoimmune diseases, retinal  
 CC neovascularisation syndromes, scarring and uterine fibroids. They may  
 CC also be useful in wound healing and in contraception. The present  
 CC sequence represents a nucleic acid sequence of the invention.  
 XX SQ Sequence 2456 BP; 519 A; 722 C; 571 G; 644 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 19; DB 11; Length 2456;  
 Best Local Similarity 100.0%; Pred. No. 56;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 CTTTACTTCATAGTCTTTG 19  
 Db 1312 CTTTACTTCATAGTCTTTG 1294  
 RESULT 28  
 ADI28525/c  
 ID ADI28525 standard; cDNA; 2456 BP.  
 XX AC ADI28525;  
 XX DT 22-APR-2004 (first entry)  
 XX DE Human GPCR retinoic acid induced 3 (RAI-3) cDNA.  
 XX KW Retinoic acid induced 3; RAI-3; human; G-protein coupled receptor; GPCR;  
 KW antiinflammatory; immunosuppressive; cytostatic; cardiant; antiallergic;  
 KW broncholytic; gene therapy; gene; single nucleotide polymorphism; SNP;  
 KW chromosome 12p13-p12.3; ss.  
 XX OS Homo sapiens.  
 PH Key Location/Qualifiers  
 FT variation /tag= a  
 FT /label= RAI-3-s1  
 FT /note= "located in 5' untranslated region"  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT CDS 254..1327  
 FT /\*tag= b  
 FT /product= "Human RAI3"  
 FT variation replace(364,y)  
 FT /\*tag= c  
 FT /label= RAI-3-s2  
 FT /note= "exon 1, silent (Ala/Ala)"  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT variation replace(511,y)  
 FT /\*tag= d  
 FT /label= RAI-3-s3  
 FT /note= "exon 2, silent (Ile/Ile)"  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT variation replace(523,y)  
 FT /\*tag= e  
 FT /label= RAI-3-s4  
 FT /note= "exon 2, silent (Asp/Asp)"  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT variation replace(605,r)  
 FT /\*tag= f  
 FT /label= RAI-3-s6  
 FT /note= "exon 2, missense (Ser/Gly)"  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT variation replace(797,r)

```
FT FT /*tag= g
FT FT /label= RAI-3-s5
FT FT /note= "exon 2, missense (Thr/Ala)"
FT FT /standard_name= "Single nucleotide polymorphism"
FT FT
FT FT variation
FT FT /*tag= h
FT FT /label= RAI-3-s8
FT FT /note= "silent (Pro/Pro)"
FT FT /standard_name= "Single nucleotide polymorphism"
FT FT
FT FT variation
FT FT replace(1173,r)
FT FT /*tag= i
FT FT /label= RAI-3-s9
FT FT /note= "missense (Gln/Arg)"
FT FT /standard_name= "Single nucleotide polymorphism"
FT FT
XX PN W02004001060-A2.
XX XX
XX PD 31-DEC-2003.
XX XX
XX XX 20-JUN-2003; 2003WO-US019255.
XX PF
XX PR 20-JUN-2002; 2002US-0390850P.
XX PR 29-AUG-2002; 2002US-0407006P.
XX XX
XX PA (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX XX
XX XX Whitney GS, Opitck G, Garulacan L, Ramanathan CS, McKinnon M;
XX PI Bennett KL, Barber LE, Cacace A, Tsuchihaehi Z;
XX XX
XX DR WPI; 2004-090973/09.
XX DR P-PSDB; ADI28526.
XX XX
XX PT New nucleic acid molecule encoding a human G-protein coupled receptor,
XX PT RAI3, useful for preventing, treating or ameliorating chronic obstructive
XX PT pulmonary disease (COPD), COPD-like disorder, or the underlying symptoms
XX PT of COPD.
XX PS
XX PS Claim 21; SEQ ID NO 18; 301pp; English.
XX XX
XX CC The present sequence is that of cDNA encoding a human G-protein coupled
XX CC receptor (Class C, Group 5), denoted retinoic acid induced 3 (RAI-3).
XX CC Proteomics methods were used to isolate cigarette smoke-inducible
XX CC tyrosine phosphorylated proteins from airway epithelial cells. RAI-3 was
XX CC identified as being tyrosine phosphorylated and/or as being
XX CC associated/complexed with tyrosine phosphorylated proteins only in those
XX CC cells that had been exposed to cigarette smoke. Since RAI-3 is primarily
XX CC expressed in lung tissue, and since cigarette smoke is a major causative
XX CC factor of chronic obstructive pulmonary disease (COPD), RAI-3 provides a
XX CC novel cellular target for identifying modulators, e.g. agonists or
XX CC antagonists, useful for the treatment and/or prevention of COPD and
XX CC related disorders such as emphysema and chronic bronchitis. RAI-3
XX CC modulators, e.g. agonists and antagonists, especially antisen-
XX CC compounds, can be used to treat COPD and other disorders and diseases
XX CC associated with regulation of NF-kB and/or its associated or interacting
XX CC signaling molecules. Single nucleotide polymorphisms (SNPs) detected in
XX CC the RAI-3 gene are useful for determining COPD association in
XX CC individuals. RAI-3 nucleic acid molecules and polypeptides are useful for
XX CC preventing, treating or ameliorating disorders related to aberrant GPCR
XX CC signaling or cell cycle regulation, pulmonary disorders, inflammatory
XX CC lung disorders, COPD, the underlying symptoms of COPD, COPD-related
XX CC disorders or conditions, autoimmune disorders, disorders related to
XX CC hyperimmune activity, inflammatory conditions, disorders related to
XX CC aberrant acute phase responses, hypercongenital conditions, birth
XX CC defects, necrotic lesions, wounds, organ transplant rejection, renal
XX CC diseases, ischaemia-reperfusion injury, heart disorders, disorders
XX CC related to aberrant signal transduction, proliferation disorders,
XX CC cancers, HIV infection, asthma, cystic fibrosis, pulmonary fibrosis,
XX CC ulcerative colitis, cerebral infarct, myocardial infarct, diabetic
XX CC nephropathy, allergic rhinitis, Crohn's disease, atherosclerosis,
XX CC rheumatoid arthritis, inflammatory/autoimmune disorders, glioblastoma,
XX CC pulmonary small cell undifferentiated carcinoma, carcinoma of the breast,
XX CC colon, lung, ovary, pancreas, prostate, non-Hodgkin's lymphoma, disorders
XX CC associated with aberrant cell adhesion, I-CAM function and/or regulation,
```

```
CC E-selectin function and/or regulation, or aberrant NF-kB function and/or
CC regulation (all claimed).
XX
XX SQ Sequence 2456 BP; 516 A; 719 C; 570 G; 643 T; 0 U; 8 Other;
Query Match 100.0%; Score 19; DB 12; Length 2456;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CTTTACTTCATAGTCTTTG 19
|||||
Db 1312 CTTTACTTCATAGTCTTTG 1294
|||||
RESULT 29
ADI28459/c
ID ADI28459 standard; cDNA; 2456 BP.
XX
XX AC ADI28459;
XX DT
XX DT 22-APR-2004 (first entry)
XX DE Human GPCR retinoic acid induced 3 (RAI-3) cDNA.
XX KW Retinoic acid induced 3; RAI-3; human; G-protein coupled receptor; GPCR;
KW antiinflammatory; immunosuppressive; cytostatic; cardiant; antiallergic;
KW broncholytic; gene therapy; gene; single nucleotide polymorphism; SNP;
KW chromosome 12p13-pl2.3; ss.
XX
XX OS Homo sapiens.
XX FH Key
XX FT Location/Qualifiers
XX FT replace(112,a)
XX FT /*tag= a
XX FT /label= RAI-3-s1
XX FT /note= "located in 5' untranslated region"
XX FT /standard_name= "Single nucleotide polymorphism"
XX FT
XX FT CDS
XX FT 254..1327
XX FT /*tag= b
XX FT /product= "Human RAI3"
XX FT replace(364,t)
XX FT /*tag= c
XX FT /label= RAI-3-s2
XX FT /note= "exon 1, silent (Ala/Ala)"
XX FT /standard_name= "Single nucleotide polymorphism"
XX FT
XX FT variation
XX FT replace(511,t)
XX FT /*tag= d
XX FT /label= RAI-3-s3
XX FT /note= "exon 2, silent (Ile/Ile)"
XX FT /standard_name= "Single nucleotide polymorphism"
XX FT
XX FT variation
XX FT replace(523,t)
XX FT /*tag= e
XX FT /label= RAI-3-s4
XX FT /note= "exon 2, silent (Asp/Asp)"
XX FT /standard_name= "Single nucleotide polymorphism"
XX FT
XX FT variation
XX FT replace(605,g)
XX FT /*tag= f
XX FT /label= RAI-3-s6
XX FT /note= "exon 2, missense (Ser/Gly)"
XX FT /standard_name= "Single nucleotide polymorphism"
XX FT
XX FT variation
XX FT replace(797,g)
XX FT /*tag= g
XX FT /label= RAI-3-s5
XX FT /note= "exon 2, missense (Thr/Ala)"
XX FT /standard_name= "Single nucleotide polymorphism"
XX FT
XX FT variation
XX FT replace(111,g)
XX FT /*tag= h
XX FT /label= RAI-3-s8
XX FT /note= "silent (Pro/Pro)"
XX FT /standard_name= "Single nucleotide polymorphism"
XX FT
XX FT variation
XX FT replace(1173,g)
XX FT /*tag= i
XX FT /label= RAI-3-s9
```

FT /note= "missense (Gln/Arg)"  
FT /standard\_name= "Single nucleotide polymorphism"  
PN WO2004001060-A2.  
XX  
XX 31-DEC-2003.  
XX  
XX 20-JUN-2003; 2003WO-US019255.  
XX  
XX 20-JUN-2002; 2002US-0390850P.  
PR 29-AUG-2002; 2002US-0407006P.  
XX  
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.  
XX  
XX Whitney GS, Opitack G, Garulacan L, Ramanathan CS, McKinnon M;  
PI Bennett KL, Barber LE, Cacace A, Tsuchihashi Z;  
XX  
XX WPI; 2004-090973/09.  
DR P-PSDB; ADI28460.  
DR GENBANK; NM\_003979.  
XX  
XX New nucleic acid molecule encoding a human G-protein coupled receptor,  
FT RAI3, useful for preventing, treating or ameliorating chronic obstructive  
FT pulmonary disease (COPD), COPD-like disorder, or the underlying symptoms  
FT of COPD.  
XX  
XX Claim 1; SEQ ID NO 2; 301pp; English.  
XX  
XX The present sequence is that of cDNA encoding a human G-protein coupled  
CC receptor (Class C, Group 5), denoted retinoic acid induced 3 (RAI-3).  
CC Proteomic methods were used to isolate cigarette smoke-inducible  
CC tyrosine phosphorylated proteins from airway epithelial cells. RAI-3 was  
CC identified as being tyrosine phosphorylated and/or as being  
CC associated/complexed with tyrosine phosphorylated proteins only in those  
CC cells that had been exposed to cigarette smoke. Since RAI-3 is primarily  
CC expressed in lung tissue, and since cigarette smoke is a major causative  
CC factor of chronic obstructive pulmonary disease (COPD), RAI-3 provides a  
CC novel cellular target for identifying modulators, e.g. agonists or  
CC antagonists, useful for the treatment and/or prevention of COPD and  
CC related disorders such as emphysema and chronic bronchitis. RAI-3  
CC modulators, e.g. agonists and antagonists, especially antisense  
CC compounds, can be used to treat COPD and other disorders and diseases  
CC associated with regulation of NF-kB and/or its associated or interacting  
CC signaling molecules. Single nucleotide polymorphisms (SNPs) detected in  
CC the RAI-3 gene are useful for determining COPD association in  
CC individuals. RAI-3 nucleic acid molecules and polypeptides are useful for  
CC preventing, treating or ameliorating disorders related to aberrant GPCR  
CC signaling or cell cycle regulation, pulmonary disorders, inflammatory  
CC lung disorders, COPD, the underlying symptoms of COPD, COPD-related  
CC disorders or conditions, autoimmune disorders, disorders related to  
CC hyperimmune activity, inflammatory conditions, disorders related to  
CC aberrant acute phase responses, hypercongenital conditions, birth  
CC defects, necrotic lesions, wounds, organ transplant rejection, renal  
CC diseases, ischemia-reperfusion injury, heart disorders, disorders  
CC related to aberrant signal transduction, proliferation disorders,  
CC cancers, HIV infection, asthma, cystic fibrosis, pulmonary fibrosis,  
CC ulcerative colitis, cerebral infarct, myocardial infarct, diabetic  
CC nephropathy, allergic rhinitis, Crohn's disease, atherosclerosis,  
CC rheumatoid arthritis, inflammatory/autoimmune disorders, glioblastoma,  
CC pulmonary small cell undifferentiated carcinoma, carcinoma of the breast,  
CC colon, lung, ovary, pancreas, prostate, non-Hodgkin's lymphoma, disorders  
CC associated with aberrant cell adhesion, I-CAM function and/or regulation,  
CC E-selectin function and/or regulation, or aberrant NF-kB function and/or  
CC regulation (all claimed).  
XX  
XX Sequence 2456 BP; 519 A; 722 C; 571 G; 644 T; 0 U; 0 Other;  
SQ

Query Match 100.0%; Score 19; DB 12; Length 2456;  
Best Local Similarity 100.0%; Pred. No. 56;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19  
|||||

Db 1312 CTTTACTTCATAGTCTTTG 1294  
RESULT 30  
ADR48221/C  
ID ADR48221 standard; cDNA; 2456 BP.  
XX AC ADR48221;  
XX  
XX 18-NOV-2004 (first entry)  
XX Human retinoic acid induced 3 (RAI3) encoding cDNA SEQ:9.  
XX  
XX pancreatic cancer-associated transcript; pancreatic cancer; human;  
KW cytoskeletal; gene therapy; protein therapy; retinoic acid induced 3; RAI3;  
KW gene; ss.  
XX  
XX Homo sapiens.  
OS  
XX  
XX Key Location/Qualifiers  
FH CDS 254..1327  
FT /\*tag= a  
FT /product= "retinoic acid induced 3 (RAI3)"  
XX  
XX WO2004074510-A1.  
XX  
XX 02-SEP-2004.  
XX  
XX 18-FEB-2004; 2004WO-AU000194.  
XX  
XX 18-FEB-2003; 2003AU-00900747.  
XX  
XX (GARV-) GARVAN INST MEDICAL RES.  
XX  
XX Biankin A, Segara D, Henshall S, Sutherland R;  
PI WPI; 2004-635591/61.  
XX P-PSDB; ADR48222.  
XX  
XX Detecting pancreatic cancer-associated transcript in a biological sample,  
FT useful for diagnosing or treating the disease, comprises contacting the  
FT sample with a polynucleotide that selectively hybridizes to a specific  
FT sequence.  
XX  
XX Claim 7; SEQ ID NO 9; 263pp; English.  
XX  
XX The present invention describes a method for detecting a pancreatic  
CC cancer-associated transcript in a biological sample. The method comprises  
CC contacting the biological sample with a polynucleotide that selectively  
CC hybridizes to a sequence at least 80% identical to a sequence as shown in  
CC any one of Tables 3 to 25 in the specification or having the GenBank  
CC Accession Number AF279145. Also described: (1) diagnosing pancreatic  
CC cancer in a human or animal subject being tested, determining the  
CC likelihood that a subject having a pancreatic cancer will survive, or  
CC determining the suitability of a subject having a pancreatic cancer for  
CC surgical resection therapy; (2) detecting a pancreatic cancer-associated  
CC polypeptide in a biological sample; (3) determining the likelihood that a  
CC subject having a pancreatic cancer will survive; and (4) monitoring the  
CC efficacy of a therapeutic treatment of pancreatic cancer. A pancreatic  
CC cancer-associated transcript has cytostatic activity, and can be used in  
CC gene and protein therapy. A pancreatic cancer-associated transcript  
CC polynucleotide, a vector comprising the polynucleotide, an isolated  
CC polypeptide or an antibody that binds to the isolated polypeptide can be  
CC used for diagnosing or prognosing pancreatic cancer or for preparing a  
CC medicament for the treatment of pancreatic cancer. The prognostic or  
CC diagnostic methods are useful for the early detection of pancreatic  
CC cancer or its metastases, and for monitoring the progress of disease  
CC as during remission or following surgery or chemotherapy. The present  
CC sequence encodes human retinoic acid induced 3 (RAI3), which is used in  
CC the exemplification of the present invention.  
XX  
XX Sequence 2456 BP; 519 A; 722 C; 571 G; 644 T; 0 U; 0 Other;

```
Query Match      100.0%; Score 19; DB 13; Length 2456;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 1 CTTTACTTCATAGTCTTTG 19  
|||  
1312 CTTTACTTCATAGTCTTTG 1294  
Db

RESULT 31

ACN39398/c  
ID ACN39388 standard; cDNA; 2456 BP.  
XX  
XX AC ACN39388;  
XX  
XX  
DT 18-NOV-2004 (first entry)  
XX  
XX Tumour-associated antigenic target (TAT) cDNA DNA26771, SEQ ID NO:3498.  
XX  
XX Tumour-associated antigenic target; TAT; human; overexpression; cancer;  
KW tumour; diagnosis; cell proliferative disorder; breast cancer;  
KW colorectal cancer; lung cancer; ovarian cancer; liver cancer;  
KW central nervous system cancer; bladder cancer; pancreatic cancer;  
KW cervical cancer; melanoma; leukaemia; hybridisation probe;  
KW chromosome identification; chromosome mapping; gene mapping;  
KW gene therapy; cytostatic; gene; ss.

**Homo sapiens.**

WO2004030615-A2.

15-APR-2004

29-SEP-2003.

02-0000-2003: 2002115-0414971B

SECRET - CONFIDENTIAL

[illegible]

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

P-PSDB; ABM81354.

New tumor-association

proliferative disorder, e

XX The invention relates to human tumour-associated antigenic target (TAT)  
CC polypeptides, and their related nucleic acids. The TAT polypeptides are  
CC overexpressed in cancer tissues compared to normal tissues, and may thus  
CC serve as effective targets for the diagnosis and treatment of cancer in  
CC mammals. The invention also relates to nucleic acid and polypeptide  
CC sequences at least 80% identical to the TAT nucleic acids and  
CC polypeptides; expression vectors and host cells comprising a TAT nucleic  
CC acid; an antibody specific for a TAT polypeptide; a peptide or organic  
CC molecule which binds to a TAT polypeptide; fusion proteins comprising a  
CC TAT polypeptide; and methods and compositions for the treatment or  
CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,  
CC antibodies, antagonists, binding molecules and compositions are useful  
CC for diagnosing or treating a cell proliferative disorder associated with  
CC increased TAT expression, particularly cancers such as breast cancer,  
CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder  
CC cancer, pancreatic cancer, cervical cancer, cancers of the central  
CC nervous system, melanoma and leukaemia. TAT nucleic acids may further be  
CC used as hybridisation probes, in chromosome and gene mapping, in  
CC chromosome identification and in gene therapy. The present sequence  
CC represents a TAT nucleic acid of the invention

Sequence 2456 BP: 519 A; 722 C; 571 G; 644 T; 0 U; 0 Other;

Query Match	100.0%;	Score 19;	DB 13;	Length 2456;
Best Local Similarity	100.0%;	Pred. No. 56;		

Qy 1 CTTTACTTCATAGTCTTTG 19  
          |||||  
Db 1312 CTTTACTTCATAGTCTTTG 1294

RESULT 32

ADR43992/C  
ID ADR43992 standard; DNA; 2456 BP.  
XX  
AC  
XX ADR43992;  
XX  
XX 18-NOV-2004 (first entry)  
DT  
DE Human breast tumour associated gene clone-02 SEQ ID NO:2.

cancer; tumour; breast cancer; lung cancer; colon cancer; kidney cancer;  
 KW antibody; cytostatic; gene therapy; human; breast tumour tissue;  
 KW breast tissue; gene; ds.

OS Homo sapiens.

PN WO2004074506-A2.

02-SEP-2004

XX  
DE  
30-JAN-2004.

13-FEB-2003 2003US-0447800P

XX 2 (M) C 1 MED CENT 1 EN

XX

XX

XX

PT cancer, comprises detecting and blocking the over expression of a gene of a protein found in breast, lung, colon, or kidney tissue.

XX  
PS  
Claim 2: SEQ ID NO 2: 220pp: English.

The present invention describes a method for diagnosing and treating a cancer, e.g. breast, lung, colon, or kidney cancer. The method comprises detecting and blocking the over expression of a gene of a protein found in breast, lung, colon, or kidney tissue. Also described is an antibody or binding portion of an antibody that specifically binds a protein found in breast tissue, lung tissue, colon tissue, or kidney tissue. The antibody has cytostatic activity, and can be used in gene therapy. The methods, antibodies, polynucleotides and polypeptides from the present invention are useful for detecting, diagnosing, preventing and treating cancer, e.g. breast, lung, colon, or kidney cancer. The present sequence represents a nucleotide sequence given in the present invention, which is over expressed in human breast tumour tissue. N.B. All 385 sequences referenced in this patent are detailed in the US provisional application SN 60/447,900, filed 02/13/2003. In this application only one sequence, representing the longest sequence of each of the 65 clones is listed in Appendix A (and given as SEQ ID NO:1 to 65 in the Sequence Listing).

XX  
S0  
Sequence 2456 BP: 517 A: 710 C: 567 G: 643 T: 0 U: 19 Other:

Query Match 100.0% Score 19 DB 13 Length 2456

Best Local Similarity	100.0%;	Pred. No. 56;			
Matches	18.	Conservative			
0:	Mismatches	0:	Indels	0:	Gaps

Qy 1 CTTACTTCATAGTCTTTG 19  
|||||  
Db 1312 CTTACTTCATAGTCTTTG 12



```

RESULT 33
ADU06126/C
ID ADU06126 standard; DNA; 2456 BP.
XX AC ADU06126;
XX AC
XX DT 27-JAN-2005 (first entry)
XX DE Novel bronchial cancer-associated human gene SeqID350.
XX KW bronchial cancer; cytostatic; tumour-associated protein;
KW cancer detection; metastasis; tumour; gene; ds; human.
XX OS Homo sapiens.
XX DE10316701-A1.
XX PD 04-NOV-2004.
XX PF 09-APR-2003; 2003DE-01016701.
XX PR 09-APR-2003; 2003DE-01016701.
XX PA (HINZ/) HINZMANN B.
XX PA (HERM/) HERMANN K.
XX PA (CAST/) HEIDEN CASTANOS-VELEZ E.
XX PI Mennerich D, Bruemendorf T, Heiden E, Hermann K, Kinnemann H;
PI Li X, Roepcke S, Staub E, Hinzmann B, Rosenthal A, Pilarsky C;
XX WPI: 2004-786403/78.
XX P-PSDB; ADU06613.
XX PT New nucleic acid, and derived proteins, useful for diagnosis of bronchial
XX cancer and in screening for therapeutic and diagnostic agents.
XX PS Claim 1; SEQ ID NO 350; 1381bp; German.
XX CC This invention relates to a novel isolated nucleic acid associated with
XX bronchial cancer comprising 489 defined sequences given in the
XX specification. The invention may be useful for the production of
XX compounds with a cytostatic activity through the inhibition of expression
XX or activity of tumour-associated proteins. The novel DNA sequences and
XX the proteins/peptides encoded by them are used for detecting bronchial
XX cancer or determining the risk of developing it and to screen for
XX specific binding partners of the DNA or protein sequences, where the
XX binding partners are potentially useful as agents for treating or
XX diagnosing bronchial cancer. The DNA or protein sequences can also be
XX used for prognosis, detection of metastases and for secondary treatment
XX (of tumours that have been stabilised or are no longer detectable).
XX Detecting abnormal expression of the DNA sequences provides early
XX diagnosis of bronchial cancers. The present sequence is that of a novel
XX bronchial cancer-associated human gene sequence of the invention.
XX SQ Sequence 2456 BP; 519 A; 722 C; 571 G; 644 T; 0 U; 0 Other;

Query Match 100.0%; Score 19; DB 13; Length 2456;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGCTTTG 19
Db 1312 CTTTACTTCATAGCTTTG 1294

RESULT 34
AEA00088/C
ID AEA00088 standard; cDNA; 2456 BP.
XX AC AEA00088;
XX AC
XX DT 28-JUL-2005 (first entry)

```

```

XX DE Human TAT115 cDNA SEQ ID NO:40.
XX KW ss; gene; tumor-associated antigen; cytostatic; breast tumor;
KW endocrine disease; gynecology and obstetrics; neoplasm; colon tumor;
KW gastrointestinal disease; rectal tumor; endometroid carcinoma;
KW genitourinary disease; renal tumor; lung tumor; respiratory disease;
KW ovary tumor; skin tumor; liver tumor.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT CDS 254..1327
FT /*tag= a
FT /product= "TAT115"
XX US2005106644-A1.
XX PD 19-MAY-2005.
XX PF 08-SEP-2004; 2004US-00936626.
XX PR 20-JUN-2001; 2001US-0299500P.
XX PR 29-JUN-2001; 2001US-0301880P.
XX PR 18-SEP-2001; 2001US-0323268P.
XX PR 19-JUN-2002; 2002US-00177488.
XX PR 26-MAR-2004; 2004US-0557116P.
XX PR 04-AUG-2004; 2004US-0598899P.
XX PA (GETH ) GENENTECH INC.
XX PI Cairns B, Chen R, Frantz G, Hillan KJ, Koeppen H, Phillips HS;
PI Polakis P, Spencer SD, Smith V, Williams PM, Wu TD, Zhang Z;
PI Sliwkowski M;
XX WPI: 2005-384304/39.
XX P-PSDB; AEA00166.
XX PT Novel isolated antibody capable of binding to tumor-associated antigenic
XX target polypeptide, useful for treating cell proliferative disorder e.g.
XX cancer.
XX PS Claim 1; SEQ ID NO 40; 337pp; English.
XX CC The invention relates to a novel isolated antibody binding to a
XX polypeptide having at least 80% sequence identity to a polypeptide having
XX any one of 76 fully defined 182-910 amino acid tumor-associated antigenic
XX target polypeptide (TAT) sequences (AEA00127-AEA00202) given in the
XX specification, a polypeptide having any one of (AEA00127-AEA00202),
XX lacking its associated signal peptide, or an extracellular domain of a
XX polypeptide having any one of (AEA00127-AEA00202). The polypeptide is
XX encoded by the nucleotide sequence having any one of (AEA00049-AEA00126).
XX An antibody of the invention has cytostatic activity. The antibody is
XX useful for inhibiting growth of a cell expressing TAT188, which involves
XX contacting the cell with the antibody. The cell is a cancer cell chosen
XX from breast, colon, rectum, endometrium, kidney, lung, ovary, skin and
XX liver cell. The cancer cell is a mammalian cell, preferably a human cell.
XX The antibody is also useful for detecting the level of TAT188
XX polypeptide expressed in a test cell relative to a control cell, and for
XX detecting the level of TAT188 polypeptide or a polypeptide having at
XX least 80 % sequence identity to the TAT188 polypeptide sequence in a test
XX cell relative to a control cell. The antibody is useful for treating a
XX preventing a cell proliferative disorder associated with increased
XX expression or activity of a polypeptide having at least 80 % identity to
XX a TAT188 polypeptide sequence. The cell proliferative disorder is cancer.
XX The method of the invention is useful for inhibiting the growth of a
XX cancer cell. The present sequence encodes a polypeptide of the invention.
XX SQ Sequence 2456 BP; 519 A; 722 C; 571 G; 644 T; 0 U; 0 Other;

Query Match 100.0%; Score 19; DB 14; Length 2456;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```
QY      1 CTTTACTTCATAGTCCTTTG 19
      |||||
Db     1312 CTTTACTTCATAGTCCTTTG 1294

RESULT 35
AEA00608/c
ID  AEA00608 standard; cDNA; 2456 BP.
XX
AC  AEA00608;
XX
DT  28-JUL-2005 (first entry)
XX
DE  Human TAT115 cDNA sequence SeqID40.
XX
KW  antibody identification; tumor-associated antigen; cytostatic;
KW  RNA interference; gene therapy; cell death; cancer; breast tumor;
KW  colon tumor; rectal tumor; renal tumor; lung tumor; ovary tumor;
KW  skin tumor; liver tumor; gene; ss; TAT115.
XX
OS  Homo sapiens.
XX
PN  US2005107595-A1.
XX
PD  19-MAY-2005.
XX
PF  10-SEP-2004; 2004US-00938061.
XX
PR  20-JUN-2001; 2001US-0299500P.
PR  29-JUN-2001; 2001US-0301890P.
PR  18-SEP-2001; 2001US-0323268P.
PR  19-JUN-2002; 2002US-00177488.
PR  26-MAR-2004; 2004US-0557116P.
PR  04-AUG-2004; 2004US-0598899P.
XX
PA  (GETH ) GENENTECH INC.
XX
PI  Cairns B, Chen R, Frantz G, Hillan KJ, Koeppe H, Phillips HS;
PI  Polakis P, Spencer SD, Smith V, Williams PM, Wu TD, Zhang Z;
PI  Sakanaka C, Chuntharapai A, Reed CJ;
XX
DR  WPI; 2005-371577/38.
DR  P-PSDB; AEA00686.
XX
PT  Novel isolated antibody e.g. anti-E16 or anti-TAT112 antibody that binds
PT  to tumor-associated antigenic target polypeptide, useful for diagnosing
PT  or treating cancer.
XX
PS  Claim 1; SEQ ID NO 40; 96pp; English.
XX
CC  This invention relates to a novel isolated antibody, for example anti-E16
CC  or anti-TAT112 antibody, that binds to a tumor-associated antigenic
CC  target polypeptide (TAT) and that lacks an associated signal peptide
CC  sequence. The invention may be useful for the development of compounds
CC  with a cytostatic activity acting as antagonists of the TAT188
CC  polypeptide or RNA interference whilst the disclosed sequences may be
CC  useful for gene therapy. The invention is useful for inducing the death
CC  of a cell (such as a cancer cell chosen from breast, colon, rectum,
CC  endometrium, kidney, lung, ovary, skin and liver) to which it binds,
CC  inhibiting proliferation or promoting cell death of a cell expressing
CC  TAT188. In addition, the invention may be useful for detecting the level
CC  of TAT188 polypeptide in a test cell relative to a control cell, or
CC  treating or preventing a cell proliferative disorder associated with
CC  increased expression of TAT188. The novel antibody of the invention is
CC  useful for inhibiting the growth of a cancer cell and may be useful for
CC  diagnosing or treating cancer. The present sequence is that of the human
CC  TAT115 cDNA which encodes a protein against which an antibody of the
CC  invention may be targeted.
XX
SQ  Sequence 2456 BP; 519 A; 722 C; 571 G; 644 T; 0 U; 0 Other;
      Query Match      100.0%; Score 19; DB 14; Length 2456;
```

Best Local Similarity 100.0%; Pred. No. 56;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCCTTTG 19  
 |||||  
Db 1312 CTTTACTTCATAGTCCTTTG 1294

RESULT 36  
AED47480/c  
ID AED47480 standard; DNA; 2456 BP.  
XX  
AC AED47480;  
XX  
DT 15-DEC-2005 (first entry)  
XX  
DE Retinoic acid-inducible G-protein coupled receptor 3 gene, SEQ ID 2.  
XX  
KW RNA Interference; gene silencing;  
KW retinoic acid-inducible G-protein coupled receptor 3; breast tumor;  
KW cytostatic; cell growth; gene; ds.  
XX  
OS Homo sapiens.

XX  
FH Key  
CDS 254..1327  
FT /tag= a  
FT /product= "Retinoic acid-inducible G-protein coupled  
FT receptor 3"  
FT misc\_feature 338..403  
FT /tag= b  
FT /label= Transmembrane domain  
FT misc\_feature 452..514  
FT /tag= c  
FT /label= Transmembrane domain  
FT misc\_feature 497..1057  
FT /tag= d  
FT /label= 7 transmembrane receptor  
FT misc\_feature 548..601  
FT /tag= e  
FT /label= Transmembrane domain  
FT misc\_feature 656..718  
FT /tag= f  
FT /label= Transmembrane domain  
FT misc\_feature 776..832  
FT /tag= g  
FT /label= Transmembrane domain  
FT misc\_feature 887..946  
FT /tag= h  
FT /label= Transmembrane domain  
FT polyA\_signal 2403..2408  
FT /tag= i  
FT polyA\_site 2442..2456  
FT /tag= j  
XX JP2005287456-A.  
XX 20-OCT-2005.  
XX 02-APR-2004; 2004JP-00110232.  
XX 02-APR-2004; 2004JP-00110232.  
XX (NIPK ) NIPPON KAYAKU KK.  
XX (UVNI-) UNIV NIPPON IKA.  
XX Nagahata T, Emi M;  
XX WPI; 2005-737982/76.  
XX P-PSDB; AED47479.  
XX Substance that suppresses expression of retinoic acid-inducible G-protein  
PT coupled receptor-3 RAI3 gene or nucleic acid, useful for suppressing cell

PT growth, and as pharmaceutical or therapeutic agent of breast cancer.  
 PS Claim 1; SEQ ID NO 2; 22pp; Japanese.  
 XX  
 CC The invention relates to a novel substance which suppresses the  
 CC expression of a retinoic acid-inducible G-protein coupled receptor (RAI) 3  
 CC gene. The RAI3 gene suppressing agent is useful for suppressing cell  
 CC growth, and as a pharmaceutical or therapeutic agent of breast cancer,  
 CC and is also useful in identifying a cell growth inhibitory substance. The  
 CC RAI3 gene suppressing agent enables the identification of a cell growth  
 CC inhibitory substance. This polynucleotide sequence represents the  
 CC retinoic acid-inducible G-protein coupled receptor 3 gene of the  
 CC invention.  
 XX  
 SQ Sequence 2456 BP; 519 A; 722 C; 571 G; 644 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 19; DB 14; Length 2456;  
 Best Local Similarity 100.0%; Pred. No. 56;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 CTTTACTTCATAGCTTTG 19  
 Db 1312 CTTTACTTCATAGCTTTG 1294  
 RESULT 37  
 ABQ54954/c  
 ID ABQ54954 standard; cDNA; 2593 BP.  
 XX  
 AC ABQ54954;  
 XX  
 XX  
 DT 22-AUG-2002 (first entry)  
 XX  
 XX Human ovarian antigen HSKJC61 cDNA, SEQ ID NO:834.  
 XX  
 KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;  
 KW ovarian cancer; breast cancer; tumour; reproductive system disorder;  
 KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;  
 KW PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;  
 KW inflammatory condition; immune disorder; blood disorder;  
 KW cardiovascular disorder; respiratory disorder; neurological disorder;  
 KW gastrointestinal disorder; urinary system disorder; drug screening;  
 KW gene therapy; chromosome mapping; forensic analysis;  
 KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;  
 KW antiinflammatory; gynaecological; reproductive; gene; ss.  
 XX  
 XX Homo sapiens.  
 OS  
 XX  
 XX WO200200677-A1.  
 XX  
 XX 03-JAN-2002.  
 XX  
 XX 07-JUN-2001; 2001WO-US018569.  
 XX  
 XX 07-JUN-2000; 2000US-0209467P.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 XX Birse CE, Rosen CA;  
 XX  
 XX WPI; 2002-147878/19.  
 XX  
 XX P-PSDB; ABP41877.  
 XX  
 PT Isolated nucleic acid molecules encoding novel ovarian polypeptides,  
 PT useful in the prevention, treatment and diagnosis of cancer (e.g. ovarian  
 PT cancer), immune disorders, cardiovascular disorders and neurological  
 PT diseases.  
 XX  
 XX Claim 1; SEQ ID NO 834; 2922pp; English.  
 PS  
 XX  
 CC The invention relates to 2175 novel human ovarian antigens (ABP41054-  
 CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also  
 CC encompasses polypeptides 90% identical and polynucleotides 95% identical

CC to the sequences of the invention. The invention additionally relates to  
 CC recombinant vectors and host cells comprising human ovarian antigen  
 CC polynucleotides, antibodies against human ovarian antigens, and the use  
 CC of ovarian antigen polynucleotides and polypeptides in diagnosing,  
 CC treating, prognosing or preventing various ovary and/or breast-related  
 CC disorders. Such conditions include ovarian cancer and breast cancer, and  
 CC metastatic tumours of ovarian or breast origin, reproductive system  
 CC disorders (e.g., infertility, disorders of pregnancy, anovulation,  
 CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine  
 CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic  
 CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and  
 CC vaginitis), immune disorders (e.g., congenital and acquired  
 CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),  
 CC blood-related disorders (e.g., anaemia), cardiovascular disorders,  
 CC respiratory disorders, neurological disorders, gastrointestinal disorders  
 CC and urinary system disorders. Ovarian antigen polypeptides and  
 CC polynucleotides may also be used in screening for compounds which  
 CC modulate ovarian antigen expression or activity. The polynucleotides may  
 CC further be used for gene therapy, chromosome mapping, in the  
 CC identification of individuals and in forensic analysis, and the  
 CC polypeptides may be used as food additives or to prepare antibodies  
 CC useful in disease diagnosis, drug targeting and phenotyping. The present  
 CC sequence represents cDNA encoding a human ovarian antigen of the  
 CC invention. Note: The sequence data for this patent did not form part of  
 CC the printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 2593 BP; 591 A; 730 C; 605 G; 667 T; 0 U; 0 Other;

Query Match 100.0%; Score 19; DB 6; Length 2593;  
 Best Local Similarity 100.0%; Pred. No. 56;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGCTTTG 19  
 Db 1424 CTTTACTTCATAGCTTTG 1406

RESULT 38

AAH72766/c

ID AAH72766 standard; cDNA; 3371 BP.

XX AAH72766;

AC AAH72766;

DT 19-SEP-2001 (first entry)

XX Human cervical cancer marker nucleic acid 4040.  
 DE  
 XX  
 XX Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.

KW Homo sapiens.

OS

XX WO200142467-A2.

PN

XX 14-JUN-2001.

XX

PF 08-DEC-2000; 2000WO-US033312.

XX

XX 08-DEC-1999; 99US-0169681P.

XX 21-DEC-1999; 99US-0171350P.

PR 14-MAR-2000; 2000US-0189315P.

PR 12-MAY-2000; 2000US-0203791P.

PR 09-JUN-2000; 2000US-0210600P.

PR 21-JUL-2000; 2000US-0220114P.

XX

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

PA

XX Schlegel R, Deeds J, Berger A, Zhao X;

XX WPI; 2001-375006/39.

XX

XX New isolated nucleic acid for diagnosing and treating cervical cancer and

XX for assessing and detecting compounds for treating the cancer.

PT

```
XX PS Claim 1; Page 840; 1051pp; English.
XX CC The invention relates to novel genes (AAH68727-AAH73383) associated with
XX CC cervical cancer with cytostatic activity. The nucleic acids and encoded
XX CC polypeptides are useful: to assess if a patient is afflicted with
XX CC cervical cancer or has a pre-malignant condition; to monitor the
XX CC progression of cervical cancer or a premalignant condition in a patient;
XX CC and to select and/or assess the efficacy of a compound or therapy for
XX CC inhibiting cervical cancer in a patient. The nucleic acids may also be
XX CC useful for gene therapy
XX SQ Sequence 3371 BP; 741 A; 902 C; 779 G; 946 T; 0 U; 3 Other;

Query Match 100.0%; Score 19; DB 4; Length 3371;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCCTTTG 19
Db 1225 CTTTACTTCATAGTCCTTTG 1207

RESULT 39
ACN89274/c
ID ACN89274 standard; DNA; 4239 BP.
XX AC ACN89274;
XX DT 02-DEC-2004 (first entry)
XX DE Breast cancer related marker, seq id 10424.
XX KW Cancer; breast; tumour; cytostatic; marker; detection; therapy; ds.
XX OS Homo sapiens.
XX PN US2003099974-A1.
XX PN 29-MAY-2003.
XX PF 18-JUL-2002; 2002US-00198846.
XX PR 18-JUL-2001; 2001US-0306220P.
XX PA (MILL-) MILLENNIUM PHARM INC.
XX PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX WP; 2003-787014/74.
XX PT Novel isolated polypeptide associated with breast cancer, useful for
XX PT detecting presence of polypeptide in sample, as a marker for breast
XX PT cancer.
XX PS Disclosure; SEQ ID NO 10424; 36pp; English.
XX CC The invention relates to an isolated polypeptide (I) associated with
XX CC breast cancer which is encoded by a nucleic acid molecule comprising a
XX CC nucleotide sequence (S1). Further disclosed is an antibody that binds to
XX CC the polypeptide of the invention. The activity of the polypeptide of the
XX CC invention may be described as cytostatic. The antibody is useful for
XX CC detecting the presence of (I) in a sample. Nucleic acid molecules of the
XX CC invention are useful in the detection of breast tumours. (I) is useful as
XX CC a marker for breast cancer and in breast cancer therapy. Sequences given
XX CC in records ACN78851-ACN92934 represent nucleic acid markers associated
XX CC with breast cancer. Note: The sequence listing does not form part of the
XX CC specification but may be obtained in electronic format from the USPTO web
XX CC site at seqdata.uspto.gov/sequence.html?DocID=2003009974
XX SQ Sequence 4239 BP; 859 A; 1161 C; 1075 G; 1141 T; 0 U; 3 Other;

Query Match 100.0%; Score 19; DB 11; Length 4239;

Best Local Similarity 100.0%; Pred. No. 58;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCCTTTG 19
Db 1720 CTTTACTTCATAGTCCTTTG 1702

RESULT 40
ADX97494/c
ID ADX97494 standard; DNA; 6730 BP.
XX AC ADX97494;
XX DT 21-APR-2005 (first entry)
XX DE Pancreatic cancer associated human protein encoding DNA, SEQ ID 42.
XX KW pancreas tumor; cytostatic; gene; ds.
XX OS Homo sapiens.
XX PN BP1471075-A2.
XX PD 27-OCT-2004.
XX PF 31-MAR-2004; 2004EP-00090124.
XX PR 31-MAR-2003; 2003DE-01015834.
XX PA (HINZ/) HINZMANN B.
XX PA (ROSE/) ROSENTHAL A.
XX PA (PILA/) PILARSKY C.
XX PA (DAHL/) DAHL E.
XX PA (SPEC/) SPECHT T.
XX PA (LICH/) LICHTNER R.
XX PI Rosenthal A, Pilarisky C, Dahl E, Specht T, Bruemendorf T;
XX PI Lichtner R, Staub E, Roepcke S, Li X;
XX WP; 2004-768082/76.
XX P-PSDB; ADX97565.
XX PT New nucleic acid differentially expressed in pancreatic tumor tissue, for
XX PT use as diagnostic agents and in screening for therapeutic agents.
XX PS Claim 1; SEQ ID NO 42; 28pp; German.
XX CC The invention relates to a novel human nucleic acid sequence of the
XX CC pancreas and its encoded protein. The invention further comprises:
XX CC proteins and peptides, preferably isolated, that contain a sequence
XX CC encoded by the novel nucleic acid; and methods for diagnosis and
XX CC treatment of pancreatic cancer, using a substance that inhibits or binds
XX CC to the protein or its DNA, including: an antisense oligonucleotide, short
XX CC interfering RNA or ribozyme directed against the pancreatic protein, an
XX CC organic molecule, particularly having a molecular weight below 5000,
XX CC especially 300, that binds to the pancreatic DNA, an aptamer or
XX CC (monoclonal) antibody, preferably human or humanized, that binds to the
XX CC pancreatic DNA, or an anti-idiotypic antibody raised against the
XX CC monoclonal antibody, any of which may be derivatized with a reporter
XX CC group, cytotoxic compound, immunostimulant and/or radioisotope. The novel
XX CC human pancreatic proteins and their encoding DNA have cytostatic
XX CC activity. The novel sequences are useful for inhibiting transcription
XX CC and/or expression of genes and proteins associated with pancreatic
XX CC cancer. This polynucleotide sequence represents the DNA encoding one of
XX CC the novel human pancreatic proteins of the invention. Note: This sequence
XX CC is not shown in the specification, it has been electronically downloaded
XX CC from a DVD-rom provided with this specification by the European Patent
XX CC Office.
XX SQ Sequence 6730 BP; 1469 A; 1763 C; 1555 G; 1943 T; 0 U; 0 Other;

Query Match 100.0%; Score 19; DB 13; Length 6730;
```

Best Local Similarity 100.0%; Pred. No. 59;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGCTTTG 19  
|||||  
Db 1312 CTTTACTTCATAGCTTTG 1294

RESULT 41  
AAA96701  
ID AAA96701 standard; DNA; 1725 BP.  
XX  
AC AAA96701;  
XX  
XX  
DT 19-FEB-2001 (first entry)  
XX  
DE Reporter gene YGR131W from an ergosterol pathway.  
XX  
KW Reporter gene; ergosterol pathway; protein kinase C pathway;  
KW invasive growth pathway; biological pathway; ss.  
XX  
OS Saccharomyces cerevisiae.  
XX  
FH Key Location/Qualifiers  
FT CDS 1001..1525  
FT /\*tag= a  
FT /product= "ergosterol-pathway polypeptide"  
XX  
PN WO200058520-A1.  
XX  
XX  
PD 05-OCT-2000.  
XX  
XX 29-MAR-2000; 2000WO-US008555.  
XX  
PR 31-MAR-1999; 99US-00282243.  
XX  
XX (ROSE-) ROSETTA INPHARMATICS INC.  
XX  
XX Roberts CJ;  
XX  
DR WPI; 2000-664929/64.  
DR P-PSDB; AAB19091.  
XX  
XX Identifying reporter and target genes for particular biological pathway  
PT of interest, useful for drug designing, involves identifying a gene which  
PT clusters to geneset associated with biological pathway.  
XX  
PS Claim 45; Fig 8; 239pp; English.  
XX  
CC The present sequence represents a reporter gene from the ergosterol  
CC pathway. The specification also describes reporter genes from the protein  
CC kinase C pathway, and the invasive growth pathway. The genes are  
CC identified using methods of the invention. The specification describes  
CC methods for identifying a reporter gene or a target gene for a particular  
CC biological pathway in a cell. The methods comprise identifying a gene  
CC which clusters to a geneset associated with the pathway. The gene  
CC identified is necessary for normal function of the pathway. The methods  
CC are useful for analysing the activity of biological pathways and in drug  
CC designing, drug therapies, or other biological agents e.g. insecticides,  
CC herbicides, fungicides, antibodies or antivirals. The reporter gene is  
CC useful as a specific target for drugs which are designed to enhance,  
CC inhibit or modulate a particular biological pathway  
XX  
SQ Sequence 1725 BP; 469 A; 315 C; 365 G; 576 T; 0 U; 0 Other;  
Query Match 91.6%; Score 17.4; DB 3; Length 1725;  
Best Local Similarity 94.7%; Pred. No. 3e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGCTTTG 19  
|||||  
Db 386 CTTTACTTCATAGCTTTG 404

RESULT 42  
ADF70573/C  
ID ADF70573 standard; DNA; 1788 BP.  
XX  
AC ADF70573;  
XX  
XX  
DT 12-FEB-2004 (first entry)  
XX  
DE Orphan receptor ligand-related human protein gene SeqID196.  
XX  
KW ligand; orphan receptor protein; fusion protein; fluorescent protein;  
KW cell expression; green fluorescent protein; GFP; GFP-1; wild-type GFP;  
KW GFPuv; Enhanced GFP; EGFP; human; gene; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO2003071272-A1.  
XX  
PD 28-AUG-2003.  
XX  
PF 21-FEB-2003; 2003WO-JP001901.  
XX  
PR 22-FEB-2002; 2002JP-00045728.  
PR 23-JUL-2002; 2002JP-00213949.  
PR 11-OCT-2002; 2002JP-00298237.  
XX  
FA (TAKE ) TAKEDA CHEM IND LTD.  
XX  
XX Hinuma S, Fujii R, Ogi K, Komatsu H, Kawamata Y, Hosoya M;  
PI WPI; 2003-697654/66.  
XX P-PSDB; ADF70471.  
DR  
DR Transformation of cells with a fusion protein of an orphan receptor  
XX protein with a fluorescent protein useful for identification of ligands  
XX to the orphan receptor.  
XX  
PS Example 4; SEQ ID NO 196; 594pp; Japanese.  
XX  
CC This invention relates to a novel method of identifying ligands to an  
CC orphan receptor protein which comprises transforming cells with DNA  
CC encoding a fusion protein of the orphan receptor with a fluorescent  
CC protein, so that the fusion protein is expressed in the cells (or cell  
CC membranes isolated from them) and contacting the cells with the potential  
CC ligand to be tested. A suitable fluorescent protein for incorporation in  
CC the fusion protein is green fluorescent protein (GFP), for example GFP-1,  
CC wild-type GFP, GFPuv or Enhanced GFP (EGFP). The method is useful for the  
CC identification of ligands binding to an orphan receptor protein.  
XX  
SQ Sequence 1788 BP; 444 A; 473 C; 412 G; 459 T; 0 U; 0 Other;  
Query Match 91.6%; Score 17.4; DB 10; Length 1788;  
Best Local Similarity 94.7%; Pred. No. 3e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGCTTTG 19  
|||||  
Db 1059 CTTTACTTCATAGCTTTG 1041

RESULT 43  
AEF80156  
ID AEF80156 standard; DNA; 188273 BP.  
XX  
AC AEF80156;  
XX  
XX  
DT 06-APR-2006 (first entry)  
XX  
XX Cancer-associated polypeptide genomic DNA sequence hd28-002 SEQ ID NO:52.  
XX ds; gene; DNA microarray; cancer; neoplasm; cytostatic; diagnosis.  
XX

```

OS Homo sapiens.
XX
XX US2006024677-A1.
XX
XX PD 02-FEB-2006.
XX
XX PF 20-JUL-2004; 2004US-00895974.
XX
XX PR 20-JUL-2004; 2004US-00895974.
XX
XX PA (MORE/) MOREIS D W
PA (MALA/) MALANDRO M S.
PA (LAIA/) LAI A.
PA (TSEC/) TSE C.
PA (FATT/) FATTAEY A.
XX
XX PI Morris DW, Malandro MS, Lai A, Tse C, Fattaey A;
XX
XX DR WPI; 2006-135411/14.
XX
XX DR P-PSDB; AEF80158.
XX
XX PT Nucleic acid array for detecting cancer-associated (CA) nucleic acid,
XX consists of nucleic acid probes having specific contiguous nucleotides of
XX CA polynucleotide.
XX
XX PS Disclosure; SEQ ID NO 52; 264pp; English.
XX
XX CC The invention relates to a novel nucleic acid array (I) for detecting a
XX cancer-associated (CA) nucleic acid, consisting of 2 or more nucleic acid
XX probes each comprising 10 or more contiguous nucleotides of 2 or more CA
XX polynucleotide sequences, or its complement. The invention has cytostatic
XX activity. The nucleic acid array is useful for detecting a CA nucleic
XX acid. An antibody of the invention is useful for detecting the presence
XX or absence of cancer cells. A method of the invention is useful for
XX inhibiting expression of a CA gene in a cell, or for treating cancer. The
XX CA polynucleotide or polypeptide as mentioned in (I) or (II) is useful as
XX vaccine for treating cancer e.g. lymphoma or leukemia. The present
XX sequence represents a human genomic sequence encoding a CA polypeptide
XX (CAP) of the invention.
XX
XX SQ Sequence 188273 BP; 52885 A; 38491 C; 40162 G; 55953 T; 0 U; 782 Other;

Query Match 91.6%; Score 17.4; DB 15; Length 188273;
Best Local Similarity 94.7%; Pred. No. 4.1e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCTTTG 19
DB 53975 CATTACTTCATAGTCTTTG 53993

RESULT 44
AEE07706
ID AEE07706 standard; cDNA; 168 BP.
AC AEE07706;
XX
XX DT 26-JAN-2006 (first entry)
XX
XX DE Tobacco p450 cDNA SEQ ID NO: 1152.
XX
XX KW Plant breeding; plant; ss; coding sequence; cytochrome p450;
XX transgenic plant; senescence.
XX
XX OS Nicotiana tabacum.
XX
XX PN WO2005111217-A2.
XX
XX PD 24-NOV-2005.
XX
XX PF 27-APR-2005; 2005WO-US014803.
XX
XX PF 29-APR-2004; 2004US-0566235P.
XX

```

---

```

PR 03-SEP-2004; 2004US-00934944.
PR 03-SEP-2004; 2004US-0607357P.
PR 17-SEP-2004; 2004US-00943507.
PR 15-OCT-2004; 2004WO-US034065.
PR 15-OCT-2004; 2004WO-US034218.
PR 25-JAN-2005; 2005US-0646764P.
PR 24-MAR-2005; 2005US-0665097P.
PR 24-MAR-2005; 2005US-0665451P.
XX
XX PA (USSM-) US SMOKELESS TOBACCO CO.
XX
XX PI Xu D, Nielsen MT;
XX
XX DR WPI; 2005-786788/80.
XX
XX PT Producing a tobacco plant having decreased expression of a nicotine
XX demethylase gene comprises crossing a first tobacco plant with a second
XX tobacco plant and germinating the collected seed of an F1 progeny plant.
XX
XX PS Claim 32; SEQ ID NO 1152; 641pp; English.
XX
XX CC The invention relates to a breeding method for producing a tobacco plant
XX with reduced expression of a nicotine demethylase gene comprises crossing
XX a first tobacco plant with variant nicotine demethylase gene expression
XX with a second tobacco plant with at least one phenotypic trait to produce
XX an F1 progeny plant, the seed of which is collected and germinated to
XX produce a tobacco plant having reduced expression of a nicotine
XX demethylase gene. Also included are breeding a nicotine demethylase
XX deficiency trait into a tobacco plant, producing a tobacco seed,
XX developing a tobacco plant in a tobacco breeding program, a tissue
XX culture of regenerable tobacco cells obtained from the tobacco plant of
XX the invention, producing a tobacco product, a breeding method for
XX producing a tobacco plant having a modified attribute, a method of
XX breeding an attribute into a tobacco plant, a tobacco plant or its
XX components produced by the method of breeding a nicotine demethylase
XX deficiency trait into a tobacco plant, producing a tobacco seed, producing
XX a tobacco plant having a modified attribute or developing a tobacco plant
XX in a tobacco breeding program, an isolated genetic marker comprising a
XX nucleic acid sequence that is substantially identical to a nucleic acid
XX sequence given in the specification (the nucleic acids comprise isolated
XX cytochrome p450 cDNAs), an expression vector comprising the isolated nucleic acid
XX sequence, a plant or plant component comprising the isolated nucleic acid
XX sequence, a plant produced from a germinated seed of the plant, reducing
XX the expression or enzymatic activity of a constitutive, or an ethylene
XX induced or senescence induced tobacco polypeptide in a plant cell, and
XX increasing the expression or enzymatic activity of a constitutive, or an
XX ethylene or senescence induced tobacco polypeptide in a plant cell. The
XX phenotypic trait comprises disease resistance, high yield, high grade
XX index, curability, curing quality, mechanical harvestability, holding
XX ability, leaf quality, height, maturation, stalk size, or leaf number per
XX plant. The breeding method for producing a tobacco plant having decreased
XX expression of a nicotine demethylase gene is useful developing desirable
XX (non-genetically engineered) germplasm. The plant is useful in producing
XX (smokeless) tobacco products. The tobacco product is a moist or dry
XX snuff, a chewing tobacco, a cigarette product, a cigar product, a
XX cigarillo, a pipe tobacco, or bidis. The p450 cDNAs were isolated using
XX degenerate PCR primers designed against cytochrome p450 motifs. The
XX present sequence is a cytochrome p450 cDNA of the invention.
XX
XX SQ Sequence 168 BP; 44 A; 23 C; 34 G; 67 T; 0 U; 0 Other;

Query Match 86.3%; Score 16.4; DB 14; Length 168;
Best Local Similarity 94.4%; Pred. No. 7.4e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TTTACTTCATAGTCTTTG 19
DB 35 TTTACTTCATAGTCTTTG 52

RESULT 45
AEF98099
ID AEF98099 standard; cDNA; 168 BP.

```

XX AEP98099;  
XX  
XX  
DT 20-APR-2006 (first entry)  
XX  
DE Tobacco cytochrome P450 cDNA clone D73A-BF3, SEQ ID NO:1152.  
XX  
XX Plant breeding: crop improvement; secondary metabolite; genetic marker;  
KW plant; cytochrome P450; CYP; ss.  
XX Nicotiana tabacum.  
XX  
XX US2006041949-A1.  
XX  
XX 23-FEB-2006.  
XX  
XX 27-APR-2005; 2005US-00116881.  
XX  
PR 13-NOV-2001; 2001US-0337684P.  
PR 11-JAN-2002; 2002US-0347444P.  
PR 12-MAR-2002; 2002US-0363684P.  
PR 16-OCT-2002; 2002US-0418933P.  
PR 13-NOV-2002; 2002US-00293252.  
PR 10-JAN-2003; 2003US-00340861.  
PR 12-MAR-2003; 2003US-00387346.  
PR 08-JUL-2003; 2003US-0485368P.  
PR 18-SEP-2003; 2003US-0503989P.  
PR 16-OCT-2003; 2003US-00886947.  
PR 29-APR-2004; 2004US-0566235P.  
PR 03-SEP-2004; 2004US-00934944.  
PR 17-SEP-2004; 2004US-0607357P.  
PR 15-OCT-2004; 2004US-00943507.  
PR 15-OCT-2004; 2004WO-US034065.  
PR 15-OCT-2004; 2004WO-US034218.  
PR 25-JAN-2005; 2005US-0646764P.  
PR 24-MAR-2005; 2005US-0665097P.  
PR 24-MAR-2005; 2005US-0665451P.  
PR 19-APR-2005; 2005US-00110062.  
XX  
XX (USSM-) US SMOKELESS TOBACCO CO.  
XX  
XX Xu D, Nielsen MT;  
XX  
XX WPI; 2006-182895/19.  
XX  
XX New breeding method, useful for producing a tobacco plant having  
PT decreased expression of a nicotine demethylase gene comprising crossing  
PT germinating F1 progeny seed to produce the tobacco plant.  
XX  
XX Claim 63; SEQ ID NO 1152; 511pp; English.  
XX  
XX The invention relates to a breeding method for producing a tobacco plant  
XX having decreased expression of a nicotine demethylase gene. The method  
XX involves crossing a first tobacco plant having variant nicotine  
XX demethylase gene expression with a second tobacco plant containing at  
XX least one phenotypic trait (e.g., disease resistance, high yield etc.) to  
XX produce an F1 progeny plant; collecting the seed of the F1 progeny; and  
XX germinating the seed to produce a tobacco plant having decreased nicotine  
XX demethylase expression. The invention also relates to a tobacco plant or  
XX its components produced using the method of the invention; a tissue  
XX culture of regenerable tobacco cells obtained from such plants; a tobacco  
XX product produced from such plants; and a method of breeding a tobacco  
XX plant with a modified attribute comprising variant expression of a  
XX cytochrome P450 polynucleotide (including the nicotine demethylase gene).  
XX The invention further relates to isolated constitutive, ethylene-induced  
XX or senescence-induced genetic markers comprising tobacco cytochrome P450  
XX nucleic acid molecules, including nicotine demethylase sequences; methods  
XX for reducing or increasing the expression or activity of polypeptides  
XX encoded by these nucleic acid molecules; expression vectors, plants or  
XX plant components comprising one of these polynucleotides; and tobacco  
XX products produced from such plants. The methods of the invention are  
XX useful for the breeding (especially marker assisted breeding) of tobacco  
XX plants with decreased nicotine demethylase expression or with altered

cytochrome P450 expression or activity. Altered expression of such enzymes can result in a change in the composition of secondary metabolites such as alkaloids (e.g., nicotine), phenylpropanoids, terpenoids, lipids, cyanogenic glycosides and glucosinolates, with effects on the flavor or aroma of plant products. They may also affect herbicide tolerance, resistance to disease or insects, quality factors related to undesirable constituents, structural traits, fiber content, leaf yield, ripening, leaf curing or storage properties. Tobacco plants of the invention in which the expression of cytochrome P450 genes is altered may have desirable traits such as altered levels of nicotine or N'-nitrosonicotine. Such plants can be used in the production of tobacco products such as moist or dry snuff, chewing tobacco, cigarettes, cigars, cigarillos, pipe tobacco, bidis or smokeless tobacco products. The present sequence represents a specifically claimed tobacco cytochrome P450 polynucleotide useful in marker assisted breeding methods. Note: The sequence data for this patent is also available in electronic format directly from the US patent office at [seqdata.uspto.gov/sequence.html?DocID=20060041949](http://seqdata.uspto.gov/sequence.html?DocID=20060041949).

Query Match 86.3%; Score 16.4; DB 15; Length 168;  
Best Local Similarity 94.4%; Pred. No. 7.4e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TTTACTTCATAGTCTTTG 19  
Db 35 TTTACTTCATAGTCTTTG 52

RESULT 46  
AAA62561/c  
ID AAA62561 standard; cDNA; 354 BP.  
XX  
AC AAA62561;  
XX  
DT 14-NOV-2000 (first entry)  
XX  
DE Human breast cancer cell cDNA sequence #377.  
XX  
XX Human; arbitrary primer; cDNA synthesis; contig sequence construction;  
KW open reading frame; ORF; low stringency; cDNA sequencing; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200031299-A2.  
XX  
XX 02-JUN-2000.  
XX  
XX 19-NOV-1999; 99WO-US027430.  
XX  
XX 20-NOV-1998; 98US-00196716.  
XX  
XX (LUDW-) LUDWIG INST CANCER RES.  
XX  
XX Simpson AJG, Dias Neto E, Brentani RR;  
XX  
XX WPI; 2000-400100/34.  
XX  
XX Method for determining open reading frames of the genome of an organism  
PT using primers at low stringency conditions, useful in the construction of  
PT contigs or constructs of sequenced nucleic acid molecules.  
XX  
XX Example 7; Page 111; 113pp; English.  
XX  
XX The present sequence is a cDNA sequence obtained using a method for  
XX determining open reading frames (ORFs) of the genome of an organism. An  
XX aliquot of cDNA from human breast cancer cells was mixed with a single,  
XX arbitrary primer, Moloney murine leukaemia virus reverse transcriptase,  
XX reverse transcriptase buffer and dNTPs. The mixture was incubated under  
XX low stringency conditions to yield single stranded cDNA. The same primer  
XX was then used to amplify the cDNA by PCR. Rather than providing  
XX nucleotide sequence information from the non-coding termini of nucleic

CC acid molecules, the method provides information on the more interesting  
 CC and relevant internal portions, such as ORFs. The method also permits the  
 CC construction of contigs of sequenced nucleic acid molecules  
 XX  
 SQ Sequence 354 BP; 118 A; 73 C; 80 G; 79 T; 0 U; 4 Other;  
 Query Match 86.3%; Score 16.4; DB 3; Length 354;  
 Best Local Similarity 94.4%; Pred. No. 7.8e+02;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 TTTACTTCAGAGCTTTG 19  
 |||||  
 Db 175 TTTACTTCAGAGCTTTG 158  
 RESULT 47  
 ABT12676/c  
 ID ABT12676 standard; DNA; 354 BP.  
 XX  
 AC ABT12676;  
 XX  
 DT 02-JAN-2003 (first entry)  
 XX  
 DE Orestes sequencing method related DNA SEQ ID No 382.  
 XX  
 KW Open reading frame; ORF; low stringency; sequencing; ORESTES; bronchitis;  
 KW identification; internal region; population study; heart disease;  
 KW Alzheimer's disease; autoimmune disease; congenital disease; human; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200274994-A2.  
 XX  
 PD 26-SEP-2002.  
 XX  
 PF 01-NOV-2001; 2001WO-US046665.  
 XX  
 PR 07-NOV-2000; 2000US-0246313P.  
 XX  
 PA (LUDW-) LUDWIG INST CANCER RES.  
 XX  
 PI Simpson AJG, Dias-Neto E, Brentani RR;  
 XX  
 DR WPI; 2002-750567/81.  
 XX  
 XX Determining ORFs of the genome by contacting mRNA with a single  
 PT oligonucleotide primer, useful for population studies that identify genes  
 PT associated with diseases such as cardiovascular, autoimmune and  
 PT congenital diseases.  
 XX  
 PS Example 7; Page 160; 166pp; English.  
 XX  
 CC The invention relates to a novel method for determining open reading  
 CC frames (ORFs) of the genome of an organism comprising contacting  
 CC messenger RNA from a cell of the organism with a single, oligonucleotide  
 CC primer at low stringency. The method is an improved ORESTES sequencing  
 CC method. The methods of the present invention are useful for sequencing  
 CC of, essentially, the entire coding component of an organism,  
 CC identification of previously unknown nucleic acid molecules, assigning  
 CC previously identified nucleotide sequences to internal regions of genes  
 CC and population studies that identify genes associated with diseases such  
 CC as heart disease, bronchitis, Alzheimer's disease, autoimmune diseases  
 CC and congenital diseases. This polynucleotide represents a DNA sequence  
 CC that is used in the novel method of the invention  
 XX  
 SQ Sequence 354 BP; 118 A; 73 C; 80 G; 79 T; 0 U; 4 Other;  
 Query Match 86.3%; Score 16.4; DB 6; Length 354;  
 Best Local Similarity 94.4%; Pred. No. 7.8e+02;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 TTTACTTCAGAGCTTTG 19  
 |||||

Db 175 TTTACTTCAGAGCTTTG 158  
 RESULT 48  
 ACD91970/c  
 ID ACD91970 standard; cDNA; 354 BP.  
 XX  
 AC ACD91970;  
 XX  
 DT 23-SEP-2003 (first entry)  
 XX  
 DE Human colon cancer cell expressed cDNA #382.  
 XX  
 KW Open reading frame detection; genome sequencing; colon cancer;  
 KW breast cancer; population genome analysis; genetic shift; cancer;  
 KW antibiotic resistance; antibiotic non-tolerance; congenital disease;  
 KW agriculture; food crop genome; resistance gene; retrovirus;  
 KW influenza virus; eukaryotic pathogen detection; trypanosome; Plasmodium;  
 KW gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US2002155438-A1.  
 XX  
 PD 24-OCT-2002.  
 XX  
 PF 27-SEP-1999; 99US-00406117.  
 XX  
 PR 20-NOV-1998; 98US-00196716.  
 XX  
 PA (SIMP/) SIMPSON A J G.  
 PA (NETO/) NETO E D.  
 PA (BREN/) BRENTANI R R.  
 XX  
 PI Simpson AJG, Neto ED, Brentani RR;  
 XX  
 DR WPI; 2003-182626/18.  
 XX  
 XX Determining open reading frames of genome of an organism e.g. a human  
 PT suffering from cancer involves use of single oligonucleotide primer at  
 PT low stringency for preparing single-stranded cDNA from mRNA of  
 PT individual.  
 XX  
 PS Example 7; Page 56; 959pp; English.  
 XX  
 CC The invention describes a method of determining open reading frames in  
 CC the genome of organism, comprising contacting mRNA from cell of organism  
 CC with a single oligonucleotide primer (I) at low stringency, preparing  
 CC single-stranded cDNA by reverse transcribing mRNA with (I), amplifying  
 CC cDNA, sequencing the product, and repeating the contacting, preparing  
 CC and amplifying steps with different primers and sequencing resulting  
 CC nucleic acids. The method is useful for: determining that a known  
 CC nucleotide sequence from a genome of an organism corresponds to a  
 CC nucleic acid molecule from a genome of an organism; and for sequencing  
 CC all or part of a genome of an organism. mRNA is obtained from mammalian  
 CC or human cell which is associated with a pathological condition e.g. a  
 CC colon cancer or breast cancer cell. The method is useful for analyses of  
 CC populations of subjects and can be used to carry out genetic analyses of  
 CC large or small populations. further, it can be used to study living  
 CC systems to determine if, e.g. there have been genetic shifts which render  
 CC an individual or population more or less likely to be afflicted with  
 CC diseases such as cancer, to determine antibiotic resistance or non-  
 CC tolerance, and so forth. The method can also be used in the study of  
 CC congenital diseases, and the risk of affliction to a fetus, as well as  
 CC the study of whether the conditions are likely to be passed to offspring  
 CC through ova or sperm. The analyses for pathological conditions can be  
 CC carried out in all animals, plants, birds, fish, etc. Using this method,  
 CC in the area of agriculture, for example the genomes of food crops can be  
 CC studied to determine if resistance genes are present, defects in plant  
 CC genomes can also be studied in this way. Similarly, the method permits  
 CC determination of the pathogens which integrate into the genome, such as  
 CC retroviruses and other integrating viruses such as influenza virus, have



CC undergone shifts or mutations, which may require different approaches to  
CC therapy. This method is also applied to eukaryotic pathogens, such as  
CC trypanosomes, different types of Plasmodium, etc. The method essentially  
CC eliminates sequencing of non-coding portions. This sequence represents a  
CC polynucleotide isolated from human colon cancer cell cDNA library  
XX  
SQ Sequence 354 BP; 118 A; 73 C; 80 G; 79 T; 0 U; 4 Other;

Query Match 86.3%; Score 16.4; DB 10; Length 354;  
Best Local Similarity 94.4%; Pred. No. 7.8e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TTTACTTCATAGTCTTTG 19  
Db 175 TTTACTTCAGAGTCTTTG 158

RESULT 49  
AAC28686/c  
ID AAC28686 standard; cDNA; 407 BP.

XX AAC28686;  
XX  
XX 06-OCT-2000 (first entry)  
XX DT  
XX DE Human secreted protein 5' EST, SEQ ID NO: 32761.

XX KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
XX KW gene therapy; chromosome mapping; ss.

XX OS Homo sapiens.  
XX PN EP1033401-A2.

XX PD 06-SEP-2000.

XX XX 21-FEB-2000; 2000EP-00200610.

XX XX 26-FEB-1999; 99US-0122487P.

XX XX (GEST ) GENSET.

XX XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX XX WPI; 2000-500381/45.

XX PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
XX PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
XX PT diagnostic, forensic, gene therapy and chromosome mapping procedures.

XX PS Claim 1; SEQ ID NO 32761; 71pp + Sequence Listing; English.

XX CC The present sequence is one of a large number of 5' ESTs derived from  
XX CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
XX CC identified within the present sequence. The 5' ESTs were prepared from  
XX CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
XX CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
XX CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
XX CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
XX CC derived from the 5' ends of mRNAs and even in those cases where longer  
XX CC cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'  
XX CC ESTs are derived from mRNAs with intact 5' ends and can therefore be used  
XX CC to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in  
XX CC diagnostic, forensic, gene therapy and chromosome mapping procedures.  
XX CC They are used to obtain upstream regulatory sequences and to design  
XX CC expression and secretion vectors

SQ Sequence 407 BP; 124 A; 78 C; 81 G; 123 T; 0 U; 1 Other;

Query Match 86.3%; Score 16.4; DB 3; Length 407;  
Best Local Similarity 94.4%; Pred. No. 7.9e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TTTACTTCATAGTCTTTG 19  
Db 40 TTTACTTCATAGTCTTTG 23

RESULT 50

AAK74755/c

ID AAK74755 standard; DNA; 429 BP.

XX AAK74755;

XX AC AAK74755;

XX DT 07-NOV-2001 (first entry)

XX DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:29567.

XX KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;

XX KW cytostatic; gene therapy; vaccine; metastasis; ds.

XX OS Homo sapiens.

XX PN WO200157182-A2.

XX PD 09-AUG-2001.

XX XX 17-JAN-2001; 2001WO-US001354.

XX PR 31-JAN-2000; 2000US-0179065P.

XX PR 04-FEB-2000; 2000US-0180628P.

XX PR 24-FEB-2000; 2000US-0184664P.

XX PR 02-MAR-2000; 2000US-0186350P.

XX PR 16-MAR-2000; 2000US-0189874P.

XX PR 17-MAR-2000; 2000US-0190076P.

XX PR 18-APR-2000; 2000US-0198123P.

XX PR 19-MAY-2000; 2000US-0205515P.

XX PR 07-JUN-2000; 2000US-0209467P.

XX PR 28-JUN-2000; 2000US-0214886P.

XX PR 30-JUN-2000; 2000US-0215135P.

XX PR 07-JUL-2000; 2000US-0216647P.

XX PR 11-JUL-2000; 2000US-0216880P.

XX PR 14-JUL-2000; 2000US-0217496P.

XX PR 26-JUL-2000; 2000US-0220963P.

XX PR 14-AUG-2000; 2000US-0224518P.

XX PR 14-AUG-2000; 2000US-0224519P.

XX PR 14-AUG-2000; 2000US-0225213P.

XX PR 14-AUG-2000; 2000US-0225214P.

XX PR 14-AUG-2000; 2000US-0225266P.

XX PR 14-AUG-2000; 2000US-0225267P.

XX PR 14-AUG-2000; 2000US-0225268P.

XX PR 14-AUG-2000; 2000US-0225270P.

XX PR 14-AUG-2000; 2000US-0225447P.

XX PR 14-AUG-2000; 2000US-0225757P.

XX PR 14-AUG-2000; 2000US-0225758P.

XX PR 18-AUG-2000; 2000US-0225759P.

XX PR 22-AUG-2000; 2000US-0226681P.

XX PR 22-AUG-2000; 2000US-0226686P.

XX PR 23-AUG-2000; 2000US-0227009P.

XX PR 30-AUG-2000; 2000US-0228924P.

XX PR 01-SEP-2000; 2000US-0229287P.

XX PR 01-SEP-2000; 2000US-0229343P.

XX PR 01-SEP-2000; 2000US-0229344P.

XX PR 05-SEP-2000; 2000US-0229509P.

XX PR 05-SEP-2000; 2000US-0229513P.

XX PR 06-SEP-2000; 2000US-0230437P.

XX PR 06-SEP-2000; 2000US-0230438P.

XX PR 08-SEP-2000; 2000US-0231242P.

XX PR 08-SEP-2000; 2000US-0231243P.

XX PR 08-SEP-2000; 2000US-0231244P.



PN WO200155314-A2.  
XX 02-AUG-2001.  
XX 17-JAN-2001; 2001WO-US001334.  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-01981123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 18-AUG-2000; 2000US-0225759P.  
PR 22-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 23-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 08-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 02-OCT-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 13-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 20-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241877P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 17-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 06-DEC-2000; 2000US-0256719P.  
PR 08-DEC-2000; 2000US-025179P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 11-DEC-2000; 2000US-0251990P.  
PR 05-JAN-2001; 2001US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA Rosen CA, Barash SC, Ruben SM;  
XX  
PI

```
XX
DR WPI; 2001-502630/55.
DR P-PSDB; AAM92695.
XX
PT Polynucleotides encoding digestive system antigens, useful for
PT diagnosing, treating, preventing and/or prognosing disorders of the
PT digestive system, particularly cancer and cancer metastases.
XX
PS Claim 1; SEQ ID NO 784; 986pp; English.
XX
XX The present invention provides the protein and coding sequences of a
CC number of human digestive system antigens. These can be used in the
CC diagnosis, treatment and prevention of digestive system disorders,
CC including cancer, Meckel's diverticulum, bacterial or parasitic
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC ulcerative colitis. The present sequence is a cDNA encoding a digestive
CC system antigen of the invention
XX
SQ Sequence 491 BP; 144 A; 90 C; 68 G; 186 T; 0 U; 3 Other;

Query Match      84.2%; Score 16; DB 4; Length 491;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3 TTACTTCATAGTCTTT 18
Db      312 TTACTTCATAGTCTTT 327

RESULT 52
AAI97742
ID AAI97742 standard; cDNA; 910 BP.
XX
AC AAI97742;
XX
DT 13-NOV-2001 (first entry)
XX
DE Human neuroblastoma expressed polynucleotide SEQ ID NO 3817.
XX
DE Human neuroblastoma; malignancy; cancer; tumour marker; N-myc; TrkA; ss.
XX
OS Homo sapiens.
XX
PN WO200166719-A1.
XX
PD 13-SEP-2001.
XX
PF 02-MAR-2001; 2001WO-JP001629.
XX
PR 07-MAR-2000; 2000JP-00159195.
XX
XX (CHIB-) CHIBA PREFECTURE.
XX (HISM ) HISAMITSU PHARM CO LTD.
XX Nakagawara A;
XX WPI; 2001-565584/63.
XX
XX Nucleic acids originating in gene expressed in human neuroblastoma,
XX useful as probe or primer in diagnosing prognosis of human neuroblastoma,
XX malignancy and susceptibility indicator or tumor marker for anti-cancer
XX agents.
XX
XX Claim 1; Page 2803; 2979pp; Japanese.
XX
XX The invention relates to novel genes (AAI93926-AAI97963) expressed in
XX human neuroblastoma. The nucleic acids are applicable as a probe or
XX primer in diagnosing the prognosis of human neuroblastoma, malignancy and
XX susceptibility indicators or tumour markers for anti-cancer agents. The
XX gene information for diagnosing prognosis is related to factors similar
XX to that for N-myc and TrkA genes
XX
XX Sequence 910 BP; 203 A; 219 C; 226 G; 232 T; 0 U; 30 Other;

Query Match      84.2%; Score 16; DB 4; Length 910;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 TTTACTTCATAGTCTT 17
Db      184 TTTACTTCATAGTCTT 199

RESULT 53
AAI97741/C
ID AAI97741 standard; cDNA; 981 BP.
XX
AC AAI97741;
XX
DT 13-NOV-2001 (first entry)
XX
DE Human neuroblastoma expressed polynucleotide SEQ ID NO 3816.
XX
DE Human neuroblastoma; malignancy; cancer; tumour marker; N-myc; TrkA; ss.
XX
OS Homo sapiens.
XX
PN WO200166719-A1.
XX
PD 13-SEP-2001.
XX
PF 02-MAR-2001; 2001WO-JP001629.
XX
PR 07-MAR-2000; 2000JP-00159195.
XX
XX (CHIB-) CHIBA PREFECTURE.
XX (HISM ) HISAMITSU PHARM CO LTD.
XX Nakagawara A;
XX WPI; 2001-565584/63.
XX
XX Nucleic acids originating in gene expressed in human neuroblastoma,
XX useful as probe or primer in diagnosing prognosis of human neuroblastoma,
XX malignancy and susceptibility indicator or tumor marker for anti-cancer
XX agents.
XX
XX Claim 1; Page 2802; 2979pp; Japanese.
XX
XX The invention relates to novel genes (AAI93926-AAI97963) expressed in
XX human neuroblastoma. The nucleic acids are applicable as a probe or
XX primer in diagnosing the prognosis of human neuroblastoma, malignancy and
XX susceptibility indicators or tumour markers for anti-cancer agents. The
XX gene information for diagnosing prognosis is related to factors similar
XX to that for N-myc and TrkA genes
XX
XX Sequence 981 BP; 261 A; 226 C; 203 G; 240 T; 0 U; 51 Other;

Query Match      84.2%; Score 16; DB 4; Length 981;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 TTTACTTCATAGTCTT 17
Db      298 TTTACTTCATAGTCTT 283

RESULT 54
AAK90456
ID AAK90456 standard; DNA; 1596 BP.
XX
AC AAK90456;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human digestive system antigen genomic sequence SEQ ID NO: 4032.
```

XX Human; digestive system antigen; gene therapy; cancer; appendicitis;  
KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;  
KW digestive system disorder; Meckel's diverticulum; ds.  
XX Homo sapiens.  
XX WO200155314-A2.  
XX 02-AUG-2001.  
XX 17-JAN-2001; 2001WO-US001124.  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 18-AUG-2000; 2000US-0225759P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226686P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251989P.

```
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 14-AUG-2000; 2000US-0224518P.
XX 14-AUG-2000; 2000US-0224519P.
XX 14-AUG-2000; 2000US-0225213P.
PA (HUMA-) HUMAN GENOME SCI INC.
XX 14-AUG-2000; 2000US-0225214P.
XX 14-AUG-2000; 2000US-0225266P.
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-502630/55.
XX 14-AUG-2000; 2000US-0225270P.
XX 14-AUG-2000; 2000US-0225447P.
PT Polynucleotides encoding digestive system antigens, useful for
PT diagnosing, treating, preventing and/or prognosing disorders of the
PT digestive system, particularly cancer and cancer metastases.
XX 14-AUG-2000; 2000US-0225758P.
XX 14-AUG-2000; 2000US-0225759P.
PS Disclosure; SEQ ID NO 4032; 986pp; English.
XX 22-AUG-2000; 2000US-0226681P.
XX 22-AUG-2000; 2000US-0226868P.
CC The present invention provides the protein and coding sequences of a
CC number of human digestive system antigens. These can be used in the
CC diagnosis, treatment and prevention of digestive system disorders,
CC including cancer, Meckel's diverticulum, bacterial or parasitic
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC ulcerative colitis. The present sequence is a genomic DNA fragment
CC encoding a digestive system antigen of the invention
XX 01-SEP-2000; 2000US-0229344P.
SQ Sequence 1596 BP; 486 A; 304 C; 259 G; 547 T; 0 U; 0 Other;
Query Match 84.2%; Score 16; DB 4; Length 1596;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 TTACTTCATAGTCTTT 18
DB 1289 TTACTTCATAGTCTTT 1304
RESULT 55
AAK90457
ID AAK90457 standard; DNA; 1596 BP.
AC AAK90457;
XX 05-NOV-2001 (first entry)
DT Human digestive system antigen genomic sequence SEQ ID NO: 4033.
DE Human; digestive system antigen; gene therapy; cancer; appendicitis;
KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW digestive system disorder; Meckel's diverticulum; ds.
XX Homo sapiens.
XX WO200155314-A2.
XX 02-AUG-2001.
XX 17-JAN-2001; 2001WO-US001324.
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184684P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 22-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
```

PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246479P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
PA (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Baraah SC, Ruben SM;  
XX WPI; 2001-502630/55.

XX Polynucleotides encoding digestive system antigens, useful for  
PT diagnosing, treating, preventing and/or prognosing disorders of the  
PT digestive system, particularly cancer and cancer metastases.

XX PS Disclosure; SEQ ID NO 4033; 986pp; English.

XX CC The present invention provides the protein and coding sequences of a  
CC number of human digestive system antigens. These can be used in the  
CC diagnosis, treatment and prevention of digestive system disorders,  
CC including cancer, Meckel's diverticulum, bacterial or parasitic  
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or  
CC ulcerative colitis. The present sequence is a genomic DNA fragment  
CC encoding a digestive system antigen of the invention

XX SQ Sequence 1596 BP; 486 A; 304 C; 259 G; 547 T; 0 U; 0 Other;

Query Match 84.2%; Score 16; DB 4; Length 1596;  
Best Local Similarity 100.0%; Pred. No. 1.3e+03;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 TTACTTCATAGCTCTT 18  
Db 1289 TTACTTCATAGCTCTT 1304  
RESULT 56  
AEF99265  
ID AEF99265 standard; DNA; 3447 BP.  
XX AC AEF99265;  
XX 20-APR-2006 (first entry)  
XX Human EFHC1 splice variant DNA.  
XX neuron; EFHC1; epilepsy; therapeutic; pharmaceutical; cell death;  
KW apoptosis; diagnostic; neurological disease; Anticonvulsant; ds.  
XX OS Homo sapiens.  
XX WO2006019978-A2.  
XX 23-FEB-2006.  
XX 15-JUL-2005; 2005WO-US025093.  
XX 16-JUL-2004; 2004US-0588769P.  
XX (REGC) UNIV CALIFORNIA.  
PA (RIKE-) RIKEN INST PHYSICAL & CHEM RES.  
PA (NAME-) INST NACIONAL NEUROLOGIA & NEUROCIROUGIA.  
PA (UYHO-) UNIV HONDURAS NACIONAL AUTONOMA.  
PI Delgado-Escueta AV, Yamakawa K, Suzuki T, Medina-Hernandez MT;  
PI Villatela MEA;  
XX WPI; 2006-184370/19.  
XX EMBL; AL122084.

Inhibiting growth of a neuronal cell for treating epilepsy, comprises  
contacting a cell with EFHC1 or its analog, or an EFHC1 agonist.

Example 11; SEQ ID NO 8; 131pp; English.

This invention describes a novel of inhibiting growth of a neuronal cell  
and comprises contacting the cell with EFHC1 or its analog, or an EFHC1  
agonist e.g. an R-type voltage-dependent Ca<sup>2+</sup> channel. The invention also  
describes: a) a method of potentiating an epilepsy therapy in a patient  
by administering EFHC1 or its analog, or EFHC1 agonist, and administering  
an anti-epileptic drug; b) a method of identifying agents that promote  
cell death in a mammalian cell; c) a method of identifying an agent that  
modulates activity of an EFHC1 polypeptide; d) a method of identifying a  
mammalian cell or mammalian cell culture engineered to overexpress EFHC1;  
e) a method of monitoring the efficacy of an epilepsy treatment by  
detecting levels of AEF99260 or AEF99261 where an increased level of the  
EFHC1 polypeptide or EFHC1 polynucleotide or gene copy number in the  
biological sample compared to the level or copy number in a biological  
sample from the patient before the treatment is indicative of efficacious  
treatment; f) a method of diagnosing epilepsy in a mammalian subject; g)  
a vector or host cell containing a nucleic acid encoding an EFHC1  
polypeptide; h) an antibody that specifically binds to an EFHC1  
polypeptide and i) a diagnostic kit comprising a solid support  
(nitrocellulose, latex or plastic material), a detection reagent (anti-  
immunoglobulin, protein G, protein A or lectin) and a reporter group is  
chosen from radioisotopes, fluorescent groups, luminescent groups,  
enzymes, biotin and dye particles. the anti-epileptic drug promotes  
apoptosis and is chosen from carbamazepine, ethosuximide, CBZ epoxide,  
clonazepam, oxcarbazepine, sodium valproate, acetazolamide, clobazam,  
gabapentin, levetiracetam, phenobarbital, phenytoin,  
piracetam, primidone, tiagabine, topiramate, zonisamide or vigabatrin.  
The methods of the invention are used to a) inhibit growth of a neuronal

CC cell; b) potentiate an epilepsy therapy in a patient; c) potentiate  
CC treatment for juvenile myoclonic epilepsy (JMC); d) monitor the efficacy  
CC of epilepsy treatment and e) enables improvement of epilepsy treatment  
CC without causing substantial toxicity. This sequence represents a splice  
CC variant of the human BFHCl gene used in the invention.

XX Sequence 3447 BP; 1022 A; 709 C; 677 G; 1039 T; 0 U; 0 Other;

Query Match 84.2%; Score 16; DB 15; Length 3447;

Best Local Similarity 100.0%; Pred. No. 1.4e+03;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 TTACTTCATAGTCTTT 18

|||||

Db 1967 TTACTTCATAGTCTTT 1982

RESULT 57

ADJ80212/c

ID ADJ80212 standard; cDNA; 3737 BP.

XX

AC ADJ80212;

XX

DT 06-MAY-2004 (first entry)

XX

DE Novel human nucleic acid-associated protein coding sequence #30.

XX

de; gene; cytostatic; antiarteriosclerotic; cerebroprotective;

XX antiParkinsonian; anticonvulsant; anti-HIV; anti-allergic; antibacterial;

XX virucide; gene therapy; nucleic acid-associated protein; cancer;

XX atherosclerosis; stroke; Parkinson's disease; epilepsy;

XX Cushing's syndrome; AIDS; allergy; microarray element;

XX protein-protein interaction; drug-target interaction; gene expression;

XX chromosomal mapping; diagnosis.

XX

OS Homo sapiens.

XX

PN W02003038052-A2.

XX

PD 08-MAY-2003.

XX

PF 29-OCT-2002; 2002WO-US034846.

XX

PR 29-OCT-2001; 2001US-0348442P.

XX

PR 01-NOV-2001; 2001US-0335544P.

XX

PR 05-NOV-2001; 2001US-0337535P.

XX

PR 09-NOV-2001; 2001US-0344650P.

XX

PR 15-NOV-2001; 2001US-0334762P.

XX

PA (INCY-) INCYTE GENOMICS INC.

XX

XX Becha SD, Borowsky ML, Burford N, Chawla NK, Elliott VS;

XX Emerling BM, Forsythe IJ, Gietzen KJ, Gorvad AE, Griffin JA;

XX Hafalia AJA, Ison CH, Lal PG, Lee EA, Lee S, Lee SY, Marquis JP;

XX Ramkumar J, Sprague WN, Swarnakar A, Tang YT, Warren BA, Yang J;

XX Yue H, Zebardjian Y;

XX WPI; 2003-430514/40.

XX P-PSDB; ADJ80154.

XX

PT New human nucleic acid-associated protein (NAAP) and polynucleotide,  
PT useful for diagnosing, treating, and preventing disorders associated with  
PT aberrant expression of NAAP, e.g. cancer, AIDS, stroke or infection.

XX

PS Claim 5; SEQ ID NO 88; 443pp; English.

XX

CC The invention relates to novel human nucleic acid-associated proteins and  
CC genes encoding them, sequences that have at least 90-99 % identity to the  
CC sequences; or biologically active or immunogenic fragments of these. The  
CC polypeptides and polynucleotides are useful in diagnosing, treating and  
CC preventing disorders associated with aberrant expression of NAAP, such as  
CC cell proliferative (e.g. cancer or atherosclerosis), neurological (e.g.  
CC stroke, Parkinson's disease or epilepsy), developmental (e.g. Cushing's

CC syndrome), autoimmune/inflammatory (e.g. AIDS or allergies), or  
CC infections. These may also be used as elements on a microarray which may  
CC monitor or measure protein-protein interactions, drug-target  
CC interactions, and gene expression profiles. The polynucleotide may also  
CC be used in chromosomal mapping and in various diagnostic assays. These  
CC are also useful in assessing the effects of exogenous compounds on the  
CC expression of nucleic acids and amino acid sequences of NAAP, in  
CC facilitating drug discovery process, and in investigating the  
CC pathogenesis of diseases or medical conditions. This sequence corresponds  
CC to the gene encoding one of the proteins of the inventions.

XX Sequence 3737 BP; 1046 A; 1017 C; 1018 G; 656 T; 0 U; 0 Other;

Query Match 84.2%; Score 16; DB 10; Length 3737;

Best Local Similarity 100.0%; Pred. No. 1.4e+03;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TTACTTCATAGTCTT 17

|||||

Db 844 TTACTTCATAGTCTT 829

RESULT 58

AEC20314/c

ID AEC20314 standard; DNA; 3737 BP.

XX

AC AEC20314;

XX

DT 03-NOV-2005 (first entry)

XX

DE Human nucleic acid-associated protein (NAAP) coding sequence - SEQ ID 88.

XX

XX nucleic acid-associated protein; hyperproliferation; atherosclerosis;  
XX antiarteriosclerotic; hepatitis; antiinflammatory; hepatotropic;  
XX psoriasis; antipsoriatic; cancer; cytostatic; neurological disease;  
XX neuroprotective; epilepsy; anticonvulsant; cerebrovascular ischemia;  
XX cerebroprotective; vasotropic; dementia; neurotic; parkinsons disease;  
XX antiParkinsonian; multiple sclerosis; anemia; antianemic;  
XX Cushing's disease; endocrine-gen.; spina bifida; autoimmune disease;  
XX immunosuppressive; inflammation; acquired immune deficiency syndrome;  
XX anti-hiv; allergy; anti-allergic; contact dermatitis; dermatological;  
XX diabetes; antidiabetic; glomerulonephritis; nephrotropic; osteoporosis;  
XX osteopathic; osteoarthritis; antiarthritic; rheumatoid arthritis;  
XX antineumatic; bacterial infection; antibacterial; viral infection;  
XX virucide; fungal infection; fungicide; parasitic infection;  
XX antiparasitic; gene; ds.

OS Homo sapiens.

XX

PN US2005186569-A1.

XX

XX 25-AUG-2005.

XX

XX 31-MAR-2004; 2004US-00491472.

XX

XX 29-OCT-2001; 2001US-0348443P.

XX

XX 01-NOV-2001; 2001US-0335544P.

XX

XX 05-NOV-2001; 2001US-0337535P.

XX

XX 09-NOV-2001; 2001US-0344650P.

XX

XX 15-NOV-2001; 2001US-0334762P.

XX

XX 29-OCT-2002; 2002WO-US034846.

XX

XX (BECH/) BECHA S D.

XX (BORO/) BOROWSKY M L.

XX (BURF/) BURFORD N.

XX (CHAW/) CHAWLA N K.

XX (ELLI/) ELLIOTT V S.

XX (EMER/) EMERLING B M.

XX (FORS/) FORSYTHE I J.

XX (GIET/) GIETZEN K J.

XX (GORV/) GORVAD A E.

XX (GRIF/) GRIFFIN J A.

XX (HAFa/) HAFALIA A J A.



PA (ISON/) ISON C H.  
 PA (LALP/) LAL P.  
 PA (LEEE/) LEE E A.  
 PA (LEES/) LEE S.  
 PA (LEES/) LEE S Y.  
 PA (MARQ/) MARQUIS J P.  
 PA (RAMK/) RAMKUMAR J.  
 PA (SPRA/) SPRAGUE W W W.  
 PA (SWAR/) SWARNAKAR A.  
 PA (TANG/) TANG Y T.  
 PA (WARR/) WARREN B A.  
 PA (YANG/) YANG J.  
 PA (YUEH/) YUE H.  
 PA (ZEBAR/) ZEBARJADIAN Y.  
 XX  
 PI Becha SD, Borowsky ML, Burford N, Chawla NK, Elliott VS;  
 PI Emerling BM, Forsythe IJ, Gletzen KJ, Gorvad AE, Griffin JA;  
 PI Hafalia AJA, Lal P, Lee EA, Lee S, Lee SY, Marquis JP;  
 PI Ramkumar J, Sprague WW, Swarnakar A, Tang YT, Warren BA, Yang J;  
 PI Yue H, Zebbarjadian Y;  
 XX  
 DR WPI: 2005-570800/58.  
 DR P-PSDB; AEC20256.  
 XX  
 PT New isolated nucleic acid-associated proteins (NAAP) and polynucleotides,  
 PT useful for diagnosing, treating, or preventing cell proliferative, or  
 PT neurological, developmental, or autoimmune/inflammatory disorders, or  
 PT infections.  
 XX  
 PS Claim 12; SEQ ID NO 88; 278pp; English.  
 XX  
 CC The invention comprises the amino acid and coding sequences of human  
 CC nucleic acid-associated proteins (NAAP). The DNA and protein sequences of  
 CC the invention are useful for diagnosing, treating, or preventing  
 CC disorders associated with aberrant expression of NAAP. Specifically, the  
 CC disorders include cell proliferative disorders (e.g. atherosclerosis,  
 CC hepatitis, psoriasis, and cancer), neurological disorders (e.g. epilepsy,  
 CC stroke, dementia, Parkinson's disease, and multiple sclerosis),  
 CC developmental disorders (e.g. anemia, Cushing's syndrome, and spina  
 CC bifida), autoimmune/inflammatory disorders (e.g. AIDS, allergy, contact  
 CC dermatitis, diabetes, glomerulonephritis, osteoporosis, osteoarthritis,  
 CC and rheumatoid arthritis), and infections (e.g. bacterial, viral, fungal,  
 CC and parasitic infections). The present DNA sequence encodes a human NAAP  
 CC of the invention.  
 XX  
 SQ Sequence 3737 BP; 1046 A; 1017 C; 1018 G; 656 T; 0 U; 0 Other;  
 Query Match 84.2%; Score 16; DB 14; Length 3737;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+03;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 TTACTTCATAGTCTT 17  
 DB 844 TTACTTCATAGTCTT 829  
 RESULT 59  
 AAC75928  
 ID AAC75928 standard; cDNA; 3739 BP.  
 XX  
 AC AAC75928;  
 XX  
 DT 08-FEB-2001 (first entry)  
 XX  
 DE Human ORFX ORF1483 polynucleotide sequence SEQ ID NO:2965.  
 XX  
 KW Human; open reading frame; ORFX: detection; cytotostatic; hepatotropic;  
 KW vulnary; antipsoriatic; antiparkinsonian; neurotropic; neuroprotective;  
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW

KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
 KW thrombosis; contraceptive; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200058473-A2.  
 XX  
 PD 05-OCT-2000.  
 XX  
 XX 31-MAR-2000; 2000WO-US008621.  
 XX  
 PR 31-MAR-1999; 99US-0127607P.  
 PR 02-APR-1999; 99US-0127636P.  
 PR 05-APR-1999; 99US-0127728P.  
 PR 30-MAR-2000; 2000US-00540763.  
 XX  
 XX (CURA-) CURAGEN CORP.  
 FA  
 XX Shimkets RA, Leach M;  
 PI  
 XX WPI: 2000-602362/57.  
 DR P-PSDB; AAB41719.  
 DR  
 XX Novel nucleic acids and peptides derived from open reading frame X,  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease.  
 XX  
 PS Claim 5; Page 2200-2202; 5507pp; English.  
 XX  
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;  
 CC antipsoriatic; antiparkinsonian; neurotropic; neuroprotective; osteopathic;  
 CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;  
 CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;  
 CC dermatological; immunosuppressive; antiinflammatory; antibacterial;  
 CC antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The  
 CC sequences can be used for determining the presence of or predisposition  
 CC to, or preventing or treating pathological conditions associated with an  
 CC ORFX-associated disorder. The nucleic acids can be used to express ORFX  
 CC proteins in gene therapy vectors. The proteins and nucleic acids may be  
 CC used to treat cancers, proliferative disorders, neurodegenerative  
 CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,  
 CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester  
 CC storage, systemic lupus erythematosus, severe combined immunodeficiency  
 CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune  
 CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and  
 CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to  
 CC enhance coagulation; to inhibit thrombosis; and as a contraceptive  
 XX  
 SQ Sequence 3739 BP; 1078 A; 787 C; 764 G; 1110 T; 0 U; 0 Other;  
 Query Match 84.2%; Score 16; DB 3; Length 3739;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+03;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 3 TTACTTCATAGTCTT 18  
 DB 2259 TTACTTCATAGTCTT 2274  
 RESULT 60  
 AEF22147/c  
 ID AEF22147 standard; DNA; 3786 BP.  
 XX  
 AC AEF22147;  
 XX  
 DT 09-MAR-2006 (first entry)

XX DE Human basal transcription modulator splice variant DNA #12.  
XX KW Cytostatic; diagnosis; transcription factor; neoplasm; ds.  
XX OS Homo sapiens.  
XX PN WO2006005042-A2.  
XX PD 12-JAN-2006.  
XX PF 30-JUN-2005; 2005WO-US023708.  
XX PR 30-JUN-2004; 2004US-0584784P.  
XX PA (CEMI-) CEMINES INC.  
XX PI Shen D, Neuman T, Palm K;  
XX WPI; 2006-100776/10.  
XX DR Diagnosing cancer, useful for treating cancer, e.g. lung, breast,  
XX PT prostate, skin, gastrointestinal cancer, comprises determining the  
XX PT expression of at least one splice variant of each of the basal  
XX PT transcription factors.  
XX PS Disclosure; Fig 9; 131pp; English.  
XX CC The invention relates to a method of diagnosing cancer which comprises  
XX CC determining the expression of at least one splice variant of each of the  
XX CC basal transcription factors, where expression of each of the basal  
XX CC transcription factor splice variants is distinguished from expression of  
XX CC its wildtype isoform, and where the expression pattern of the basal  
XX CC transcription factor splice variants is indicative of cancer. The methods  
XX CC and bioactive agents are useful for treating cancer, e.g. lung cancer,  
XX CC gastrointestinal cancer, breast cancer, prostate cancer, skin cancer,  
XX CC sarcoma, endocrine cancer, neural cancer, bladder cancer, cervical  
XX CC cancer, renal cancer and hematopoietic cancer. The present sequence  
XX CC represents a human basal transcription modulator splice variant DNA.  
XX SQ Sequence 3786 BP; 1063 A; 1031 C; 1030 G; 662 T; 0 U; 0 Other;  
Query Match 84.2%; Score 16; DB 15; Length 3786;  
Best Local Similarity 100.0%; Pred. No. 1.4e+03;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 TTTACTTCATAGTCTT 17  
DB 811 TTTACTTCATAGTCTT 796  
RESULT 61  
ACN39480/c  
ID ACN39480 standard; CDNA; 4816 BP.  
XX AC ACN39480;  
XX DT 18-NOV-2004 (first entry)  
XX DE Tumour-associated antigenic target (TAT) cDNA DNA325691, SEQ ID NO:3645.  
XX KW Tumour-associated antigenic target; TAT; human; overexpression; cancer;  
XX KW tumour; diagnosis; cell proliferative disorder; breast cancer;  
XX KW colorectal cancer; lung cancer; ovarian cancer; liver cancer;  
XX KW central nervous system cancer; bladder cancer; pancreatic cancer;  
XX KW cervical cancer; melanoma; leukaemia; hybridisation probe;  
XX KW chromosome identification; chromosome mapping; gene mapping;  
XX KW gene therapy; cytostatic; gene; ss.  
XX OS Homo sapiens.  
XX PN WO2004030615-A2.  
XX PR 16-DEC-2003; 2003US-00737082.

PD 15-APR-2004.  
XX 29-SEP-2003; 2003WO-US028547.  
XX 02-OCT-2002; 2002US-0414971P.  
XX (GETH ) GENENTECH INC.  
XX Wu TD, Zhang Z, Zhou Y;  
XX WPI; 2004-347921/32.  
XX DR New tumor-associated antigenic target polypeptides and nucleic acids,  
XX PT useful in preparing a medicament for treating or detecting a  
XX PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or  
XX PT prostate cancer or tumor.  
XX PS Claim 1; SEQ ID NO 3645; 7273pp; English.  
XX CC The invention relates to human tumour-associated antigenic target (TAT)  
XX CC polypeptides, and their related nucleic acids. The TAT polypeptides are  
XX CC overexpressed in cancer tissues compared to normal tissues, and may thus  
XX CC serve as effective targets for the diagnosis and treatment of cancer in  
XX CC mammals. The invention also relates to nucleic acid and polypeptide  
XX CC sequences at least 80% identical to the TAT nucleic acids and  
XX CC polypeptides; expression vectors and host cells comprising a TAT nucleic  
XX CC acid; an antibody specific for a TAT polypeptide; a peptide or organic  
XX CC molecule which binds to a TAT polypeptide; fusion proteins comprising a  
XX CC TAT polypeptide; and methods and compositions for the treatment or  
XX CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,  
XX CC antibodies, antagonists, binding molecules and compositions are useful  
XX CC for diagnosing or treating a cell proliferative disorder associated with  
XX CC increased TAT expression, particularly cancers such as breast cancer,  
XX CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder  
XX CC cancer, pancreatic cancer, cervical cancer, cancers of the central  
XX CC nervous system, melanoma and leukaemia. TAT nucleic acids may further be  
XX CC used as hybridisation probes, in chromosome and gene mapping, in  
XX CC chromosome identification and in gene therapy. The present sequence  
XX CC represents a TAT nucleic acid of the invention  
XX SQ Sequence 4816 BP; 1275 A; 1285 C; 1265 G; 991 T; 0 U; 0 Other;  
Query Match 84.2%; Score 16; DB 13; Length 4816;  
Best Local Similarity 100.0%; Pred. No. 1.4e+03;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 TTTACTTCATAGTCTT 17  
DB 811 TTTACTTCATAGTCTT 796  
RESULT 62  
AEA61137  
ID AEA61137 standard; DNA; 127567 BP.  
XX AC AEA61137;  
XX DT 25-AUG-2005 (first entry)  
XX DE Human BCAS1 gene genomic sequence SEQ ID NO:47.  
XX KW DNA methylation; biomarker; cancer; gene; ds; BCAS1.  
XX OS Homo sapiens.  
XX PN US2005130172-A1.  
XX PD 16-JUN-2005.  
XX 27-JAN-2004; 2004US-00765790.  
XX PR 16-DEC-2003; 2003US-00737082.

PA	(FARB ) BAYER CORP.
XX	
PI	Beard C, Burgess C, Gannon A, Harvey J, Lechner JF, Li Z;
XX	
DR	WPI; 2005-456991/46.
DR	GEMBANK; NM_003657.
XX	
PT	Identifying nucleic acid sequences as biomarker for disease, by
PT	identifying nucleic acid sequences comprising methylated CpG site and
PT	down-regulated in diseased cells and comparing its expression level with
PT	demethylated nucleic acid.
XX	
FS	Claim 11; SEQ ID NO 47; 27pp; English.
XX	
CC	The invention relates to a method (M1) for identifying one or more
CC	nucleic acid sequences useful as a biomarker for a disease to be
CC	detected. (M1) involves identifying nucleic acid sequences comprising
CC	methylated CpG site in promoter-first exon region and that are down-
CC	regulated in diseased cells, comparing expression level of nucleic acid
CC	sequences with that of demethylated nucleic acid sequences and
CC	identifying nucleic acid sequences exhibiting increase in expression
CC	after demethylation. Also described: (1) detecting (M2) the presence or
CC	stage of a disease in a subject, which involves determining the degree of
CC	methylation of one or more CpG sites on nucleic acid sequences in a
CC	biological sample obtained from the subject, and determining the presence
CC	of, predposition to, or stage of the disease in the subject based on
CC	the degree of methylation; (2) monitoring the onset, progression, or
CC	regression of a disease in a subject; (3) determining the efficacy of a
CC	test compound for inhibiting a disease in a subject; and (4) a kit (I)
CC	useful for diagnosis, prognosis, staging, monitoring, and therapeutic
CC	treatment of a disease. (M1) is useful for identifying one or more
CC	nucleic acid sequences useful as a biomarker for a disease to be
CC	detected, where the nucleic acid sequences are useful for detecting, the
CC	presence or stage of a disease such as cancer e.g. colorectal cancer in a
CC	subject. The present sequence represents a specifically claimed human
CC	genomic sequence for use in the method of the invention. Note - The
CC	sequence data for this patent is not represented in the printed
CC	specification but was obtained in electronic format from the USPTO web
CC	site.
XX	
SQ	Sequence 127567 BP; 36913 A; 25847 C; 25973 G; 38934 T; 0 U; 0 Other;
	Query Match 84.2%; Score 16; DB 14; Length 127567;
	Best Local Similarity 100.0%; Pred.No. 1.7e+03;
	Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0
Qy	3 TTACTTCATAGTCTTT 18 
Db	110607 TTACTTCATAGTCTTT 110622
AAAC36870/c	
ID	AAAC36870 standard; DNA; 483 BP.
XX	
AC	AAAC36870;
XX	
DT	17-OCT-2000 (first entry)
XX	
DE	Arabidopsis thaliana DNA fragment SEQ ID NO: 15361.
XX	
KW	Hybridisation assay; Genetic mapping; Gene expression control;
KW	protein identification; signal transduction pathway; metabolic pathway;
KW	promoter; termination sequence; ss.
XX	
OS	Arabidopsis thaliana.
XX	
PN	EPI033405-A2.
XX	
PD	06-SEP-2000.
XX	
Pf	25-FEB-2000; 2000EP-00301439.
XX	

PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 17-JUL-1999; 99US-0144086P.  
PR 18-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144532P.  
PR 20-JUL-1999; 99US-0144532P.  
PR 20-JUL-1999; 99US-0144532P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 05-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 06-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 08-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149910P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150894P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151338P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.

PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 22-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161992P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 83.2%; Score 15.8; DB 3; Length 483;

Best Local Similarity 89.5%; Pred No. 1.5e+03;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19  
||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 446 CTTCACTCCATAGTCTTTG 428

RESULT 64

ADK16126/c

ID ADK16126 standard; DNA; 654 BP.

XX ADK16126;

AC ADK16126;

XX 06-MAY-2004 (first entry)

XX Nanoarchaeum equitans cancer-associated (CA) gene #39.

DE cancer-associated gene; CA gene; cancer; carcinoma; lymphoma; leukaemia;

XX ds; gene.

XX Nanoarchaeum equitans.

XX WO2003093434-A2.

XX 13-NOV-2003.

XX 01-MAY-2003; 2003WO-US013699.

XX 01-MAY-2002; 2002US-0377447P.

XX (DIVE-) DIVERSA CORP.

XX Stetter KO, Waters E, Kretz K, Podar M, Richardson T;

PI Noordewier M;

XX WPI; 2004-053041/05.

DR

Dr P-PSDB; ADK16127.  
 XX New recombinant cancer-associated genes, such as KCMU9, useful for  
 PT diagnosing or treating carcinoma, e.g. breast, colon, rectal, pancreatic,  
 PT cervical, or skin cancers, lymphomas, or leukemia.  
 XX  
 PS Claim 5; SEQ ID NO 78; 251pp; English.  
 XX  
 CC The invention comprises then amino acid and coding sequences of cancer-  
 CC associated (CA) genes isolated from Nanoarchaeum equitans. The invention  
 CC also comprises the Nanoarchaeum equitans genome. The DNA and protein  
 CC sequences of the invention are useful for diagnosing and treating cancer  
 CC (e.g. carcinoma, lymphoma, or leukaemia). The present DNA sequence  
 CC represents a Nanoarchaeum equitans CA gene of the invention.  
 XX  
 SQ Sequence 654 BP; 274 A; 87 C; 137 G; 156 T; 0 U; 0 Other;  
 Query Match 83.2%; Score 15.8; DB 12; Length 654;  
 Best Local Similarity 89.5%; Pred. No. 1.5e+03;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Qy 1 CTTTACTTCATAGTCTTTG 19  
 Db 84 CTTTAAATTGATAGTCTTTG 66  
 RESULT 65  
 ADN73034/C  
 ID ADN73034 standard; cDNA; 1131 BP.  
 XX  
 AC ADN73034;  
 XX  
 DT 15-JUL-2004 (first entry)  
 XX  
 DE Thale cress cDNA upregulated in E2Fa/Dpa expressing plants SeqID 929.  
 XX  
 KW gene; ss; plant; transgenic; E2Fa/Dpa transcription factor;  
 KW growth regulator; animal feed product; thale cress;  
 KW cell wall biosynthesis; nitrogen metabolism; carbon metabolism.  
 XX  
 XX Arabidopsis thaliana.  
 OS  
 PN WO2004035798-A2.  
 XX  
 PD 29-APR-2004.  
 XX  
 PF 20-OCT-2003; 2003WO-EP011658.  
 XX  
 PR 18-OCT-2002; 2002EP-00079408.  
 XX  
 PA (CROP-) CROPEDESIGN NV.  
 XX  
 PI Inze D, De Veylder L, Vlieghe K;  
 XX  
 DR WPI; 2004-348466/32.  
 DR P-PSDB; ADN73035.  
 XX  
 XX Altering plant characteristics, useful for producing plants for enzyme or  
 PT pharmaceutical production comprises modifying in a plant, expression of  
 PT one or more nucleic acids and/or modifying level or activity of one or  
 PT more proteins.  
 XX  
 PS Claim 1; SEQ ID NO 929; 134pp; English.  
 XX  
 CC This invention relates to a novel method for altering one or more plant  
 CC characteristics. Specifically, it refers to identifying genes that are up  
 CC - or down-regulated in transgenic plants overexpressing the heterodimeric  
 CC E2Fa/Dpa transcription factor of Arabidopsis and using these sequences to  
 CC alter plant characteristics accordingly. The present invention describes  
 CC generating transgenic plants for the production of growth regulators,  
 CC enzymes, therapeutics, pharmaceuticals and animal feed products, where  
 CC the altered plant characteristics are selected from increased yield or  
 CC biomass, enhanced survival capacity, stress tolerance, plant architecture

or physiology, altered endoreduplication, biochemistry, signal  
 transduction, storage lipid mobilisation and/or altered photosynthesis,  
 each relative to the corresponding wild type plants. Accordingly, these  
 sequences can also be useful as positive or negative selectable markers  
 during transformation of cells or tissues. The identified genes play a  
 role in a variety of biological processes such as DNA replication, cell  
 wall biosynthesis, nitrogen and/or carbon metabolism or they function as  
 CC transcription factors. This polynucleotide sequence is thale cress cDNA  
 CC upregulated 1.3 fold or more in plants overexpressing the E2Fa/Dpa  
 CC transcription factor, given in an exemplification of the invention.  
 XX  
 SQ Sequence 1131 BP; 363 A; 183 C; 284 G; 301 T; 0 U; 0 Other;  
 Query Match 83.2%; Score 15.8; DB 12; Length 1131;  
 Best Local Similarity 89.5%; Pred. No. 1.6e+03;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Qy 1 CTTTACTTCATAGTCTTTG 19  
 Db 484 CTTCACTCCATAGTCTTTG 466  
 RESULT 66  
 AAS25932/C  
 ID AAS25932 standard; cDNA; 1199 BP.  
 XX  
 AC AAS25932;  
 XX  
 DT 07-NOV-2001 (first entry)  
 XX  
 DE Human cDNA encoding a novel secreted protein, Seq ID 111.  
 XX  
 KW Human; immunosuppressive; antiarthritic; ss; antirheumatic; cytostatic;  
 KW cardiant; vasotropic; cerebroprotective; neurotropic; neuroprotective;  
 KW antibacterial; virucide; fungicide; ophthalmological; vulnerary;  
 KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;  
 KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;  
 KW cerebral ischaemia; angiogenesis; nervous system disorder;  
 KW Alzheimer's disease; infection; ocular disorder; corneal infection;  
 KW wound healing; epithelial cell proliferation; skin ageing; food additive;  
 KW preservative; antiproliferative.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200155322-A2.  
 XX  
 PD 02-AUG-2001.  
 XX  
 PR 17-JAN-2001; 2001WO-US001341.  
 XX  
 PR 31-JAN-2000; 2000US-0179065P.  
 PR 04-FEB-2000; 2000US-0180628P.  
 PR 24-FEB-2000; 2000US-0184664P.  
 PR 02-MAR-2000; 2000US-0186350P.  
 PR 16-MAR-2000; 2000US-0189874P.  
 PR 17-MAR-2000; 2000US-0190076P.  
 PR 18-APR-2000; 2000US-0198123P.  
 PR 19-MAY-2000; 2000US-0205515P.  
 PR 07-JUN-2000; 2000US-0209467P.  
 PR 28-JUN-2000; 2000US-0214886P.  
 PR 30-JUN-2000; 2000US-0215135P.  
 PR 07-JUL-2000; 2000US-0216647P.  
 PR 07-JUL-2000; 2000US-0216880P.  
 PR 11-JUL-2000; 2000US-0217487P.  
 PR 11-JUL-2000; 2000US-0217496P.  
 PR 14-JUL-2000; 2000US-0218290P.  
 PR 26-JUL-2000; 2000US-0220963P.  
 PR 26-JUL-2000; 2000US-0220964P.  
 PR 14-AUG-2000; 2000US-0224518P.  
 PR 14-AUG-2000; 2000US-0224519P.  
 PR 14-AUG-2000; 2000US-0225213P.  
 PR 14-AUG-2000; 2000US-0225214P.  
 PR 14-AUG-2000; 2000US-0225266P.

PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226686P.  
PR 22-AUG-2000; 2000US-0227009P.  
PR 23-AUG-2000; 2000US-0228924P.  
PR 30-AUG-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 14-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0246417P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.

PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256179P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-488783/53.

P-PSDB; AAU15945.

PT New nucleic acid molecules encoding 461 human secreted proteins for  
diagnosing, preventing, treating or ameliorating medical conditions and  
used as food additives or preservatives.

XX Claim 1; SEQ ID NO 111; 980pp; English.

XX The invention relates to isolated nucleic acid molecules and their  
CC encoded secreted proteins. The nucleic acids and proteins are used to  
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used  
CC in diagnosing a pathological condition or susceptibility to a  
CC pathological condition. Antibodies to the proteins can also be used in  
CC alleviating symptoms associated with the disorders and in diagnostic  
CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays  
CC (ELISA). Disorders which are diagnosed or treated include autoimmune  
CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.  
CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac  
CC arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiodenesis,  
CC nervous system disorders e.g. Alzheimer's disease, infections caused by  
CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection,  
CC and many other disorders listed in the specification. The polypeptides  
CC can also be used to aid wound healing and epithelial cell proliferation,  
CC to prevent skin aging due to sunburn, to maintain organs before  
CC transplantation, for supporting cell culture of primary tissues, to  
CC regenerate tissues and in chemotaxis. The polypeptides can also be used

CC as a food additive or preservative to increase or decrease storage  
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
CC minerals, cofactors and other nutritional components. The present  
CC sequence encodes a novel secreted protein of the invention. Note: The  
CC sequence data for this patent did not form part of the printed

Query Match 83.2%; Score 15.8; DB 4; Length 1199;  
Best Local Similarity 89.5%; Pred. No. 1.6e+03;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGCTTTG 19  
||||| ||||| |||||  
Db 692 CTTTATTTCATAGCCTTTG 674

RESULT 67  
ADB81647  
ID ADB81647 standard; DNA; 1258 BP.

XX AC ADB81647;

XX DT 04-DEC-2003 (first entry)

XX OS Human ovarian specific gene SEQ ID NO:20.

XX DE ds; gene; human; ovarian specific gene; cytostatic; ovarian cancer;  
XX KW ovary; vaccine; gene therapy; OSG; 5p25.3b.

XX XX Homo sapiens.

XX PN WO2003020897-A2.

XX XX 13-MAR-2003.

XX XX 29-AUG-2002; 2002WO-US027727.

XX XX 31-AUG-2001; 2001US-0316307P.

XX PA (DIAD-) DIADEXUS INC.

XX PI Sun Y, Liu C, Salceda S;

XX XX WPI; 2003-290181/28.

XX DR P-PSDB; ADB81590.

XX PT New ovarian specific nucleic acid, useful as a vaccine, in gene therapy,  
XX or for identifying, diagnosing, monitoring, staging, imaging and treating  
XX ovarian cancer and non-cancerous disease states of the ovary.

XX PS Example 1; Page 162-163; 220pp; English.

XX CC The invention relates to a novel isolated ovarian specific nucleic acid.  
XX CC A nucleic acid sequence of the invention has cytostatic activity. The  
XX CC ovarian specific nucleic acid or a polypeptide encoded by it is useful  
XX CC for identifying, diagnosing, monitoring, staging, imaging and treating  
XX CC ovarian cancer and non-cancerous disease states of the ovary. The ovarian  
XX CC specific nucleic acid molecule or polypeptide is also useful as a  
XX CC vaccine, in gene therapy, for producing transgenic animals and cells, or  
XX CC producing engineered ovarian tissue for treatment and research. The  
XX CC antibody that specifically binds to the polypeptide is useful for  
XX CC treating a patient with ovarian cancer, particularly for inducing an  
XX CC immune response against the ovarian cancer cell expressing the nucleic  
XX CC acid molecule or polypeptide. The sequences shown in ADB81628-ADB81688  
XX CC represent human ovarian specific genes (OSG's) of the invention.

XX SQ Sequence 1258 BP; 460 A; 200 C; 215 G; 382 T; 0 U; 1 Other;

Query Match 83.2%; Score 15.8; DB 10; Length 1258;  
Best Local Similarity 89.5%; Pred. No. 1.6e+03;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGCTTTG 19  
||||| ||||| |||||

Db 1232 CTTTATTTCATATCTTTG 1250

RESULT 68

ABX73273/c  
ID ABX73273 standard; DNA; 1259 BP.

XX AC ABX73273;

XX DT 18-MAR-2003 (first entry)

XX DE Human novel polynucleotide #101.

XX OS Homo sapiens.  
XX PN US2002132753-A1.  
XX PD 19-SEP-2002.  
XX PF 17-JAN-2001; 2001US-00764864.  
XX XX 31-JAN-2000; 2000US-0179065P.  
XX PR 04-FEB-2000; 2000US-0180628P.  
XX PR 28-JUN-2000; 2000US-0214886P.  
XX PR 07-JUL-2000; 2000US-0216647P.  
XX PR 07-JUL-2000; 2000US-0216880P.  
XX PR 11-JUL-2000; 2000US-0217487P.  
XX PR 11-JUL-2000; 2000US-0217496P.  
XX PR 14-JUL-2000; 2000US-0218290P.  
XX PR 26-JUL-2000; 2000US-0220963P.  
XX PR 26-JUL-2000; 2000US-0220964P.  
XX PR 14-AUG-2000; 2000US-0224518P.  
XX PR 14-AUG-2000; 2000US-0224519P.  
XX PR 14-AUG-2000; 2000US-0225267P.  
XX PR 14-AUG-2000; 2000US-0225268P.  
XX PR 14-AUG-2000; 2000US-0225270P.  
XX PR 14-AUG-2000; 2000US-0225447P.  
XX PR 14-AUG-2000; 2000US-0225757P.  
XX PR 14-AUG-2000; 2000US-0225758P.  
XX PR 22-AUG-2000; 2000US-0226868P.  
XX PR 30-AUG-2000; 2000US-0228924P.  
XX PR 01-SEP-2000; 2000US-0229287P.  
XX PR 01-SEP-2000; 2000US-0229343P.  
XX PR 01-SEP-2000; 2000US-0229344P.  
XX PR 01-SEP-2000; 2000US-0229345P.  
XX PR 05-SEP-2000; 2000US-0229509P.  
XX PR 08-SEP-2000; 2000US-0229513P.  
XX PR 08-SEP-2000; 2000US-0231413P.  
XX PR 21-SEP-2000; 2000US-0234223P.  
XX PR 21-SEP-2000; 2000US-0234274P.  
XX PR 25-SEP-2000; 2000US-0234977P.  
XX PR 27-SEP-2000; 2000US-0235834P.  
XX PR 29-SEP-2000; 2000US-0236327P.  
XX PR 29-SEP-2000; 2000US-0236367P.  
XX PR 29-SEP-2000; 2000US-0236368P.  
XX PR 29-SEP-2000; 2000US-0236369P.  
XX PR 29-SEP-2000; 2000US-0236370P.  
XX PR 02-OCT-2000; 2000US-0236802P.  
XX PR 02-OCT-2000; 2000US-0237037P.  
XX PR 02-OCT-2000; 2000US-0237038P.  
XX PR 02-OCT-2000; 2000US-0237039P.  
XX PR 13-OCT-2000; 2000US-0237040P.  
XX PR 20-OCT-2000; 2000US-0240960P.  
XX PR 20-OCT-2000; 2000US-0241785P.





PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135112P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140333P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140931P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142300P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0132048P.  
PR 21-JUL-1999; 99US-0132407P.  
PR 21-JUL-1999; 99US-0132484P.  
PR 22-JUL-1999; 99US-0132485P.  
PR 22-JUL-1999; 99US-0132486P.  
PR 22-JUL-1999; 99US-0132487P.  
PR 22-JUL-1999; 99US-0132863P.  
PR 23-JUL-1999; 99US-0134256P.  
PR 23-JUL-1999; 99US-0134218P.  
PR 23-JUL-1999; 99US-0134219P.  
PR 23-JUL-1999; 99US-0134221P.  
PR 26-JUL-1999; 99US-01345913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149910P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151338P.  
PR 07-SEP-1999; 99US-0151930P.  
PR 10-SEP-1999; 99US-0152363P.  
PR 13-SEP-1999; 99US-0153707P.  
PR 15-SEP-1999; 99US-0153758P.  
PR 16-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.

```
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161933P.
PR 29-OCT-1999; 99US-0162142P.

Query Match      83.2%; Score 15.8; DB 3; Length 1394;
Best Local Similarity 89.5%; Pred. No. 1.6e+03;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 CTTTACTTCATAGCTTTG 19
      ||| ||| ||| ||| ||| |||
Db      552 CTTCACTCCATAGCTTTG 534

RESULT 71
AAC33429/c
ID AAC33429 standard; DNA; 1398 BP.
XX
AC AAC33429;
XX
DT 17-OCT-2000 (first entry)
DE
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 3005.
XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic pathway;
KW promoter; termination sequence; ss.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-00301439.
XX
XX 25-FEB-1999; 99US-0121825P.
XX 05-MAR-1999; 99US-0123180P.
XX 09-MAR-1999; 99US-0123548P.
XX 23-MAR-1999; 99US-0125788P.
XX 25-MAR-1999; 99US-0126264P.
XX 29-MAR-1999; 99US-0126785P.
XX 01-APR-1999; 99US-0127462P.
XX 06-APR-1999; 99US-0128234P.
XX 08-APR-1999; 99US-0128714P.
XX 16-APR-1999; 99US-0129845P.
XX 19-APR-1999; 99US-0130077P.
XX 21-APR-1999; 99US-0130449P.
XX 23-APR-1999; 99US-0130510P.
XX 28-APR-1999; 99US-0130891P.
XX 30-APR-1999; 99US-0131449P.
XX 30-APR-1999; 99US-0132048P.
XX 04-MAY-1999; 99US-0132407P.
XX 05-MAY-1999; 99US-0132484P.
XX 06-MAY-1999; 99US-0132485P.
XX 06-MAY-1999; 99US-0132486P.
XX 07-MAY-1999; 99US-0132487P.
XX 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 27-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140921P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 21-JUL-1999; 99US-0144894P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
```

PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146336P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146398P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 26-AUG-1999; 99US-0151085P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.

PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161040P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161922P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 83.2%; Score 15.8; DB 3; Length 1398;

Best Local Similarity 89.5%; Pred. No. 1.6e+03;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 CTTTACTTCATAGTCTTTG 19  
||| ||| ||| ||| ||| |||  
Db 555 CTTCACTCCATAGTCTTTG 537

# RESULT 72

AAI60916/c

ID AAI60916 standard; cDNA; 1401 BP.

XX AC AAI60916;

XX DT 22-OCT-2001 (first entry)

XX DE Human polynucleotide SEQ ID NO 4905.

XX KW Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;

KW Peripheral nervous system; neuropathy; central nervous system; CNS;

KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;

KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;

KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;

KW leukaemia; ss.

XX OS Homo sapiens.

XX XX WO200153312-A1.

XX PD 26-JUL-2001.

XX PF 26-DEC-2000; 2000WO-US034263.

XX PR 23-DEC-1999; 99US-00471275.

PR 21-JAN-2000; 2000US-0048725.

PR 25-APR-2000; 2000US-00552317.

PR 20-JUN-2000; 2000US-00598042.

PR 19-JUL-2000; 2000US-00620312.

PR 03-AUG-2000; 2000US-00653450.

PR 14-SEP-2000; 2000US-00662191.

PR 19-OCT-2000; 2000US-00693036.

PR 29-NOV-2000; 2000US-00727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;

PI Zhou P, Goodrich R, Drmanac RT;

XX WPI; 2001-442253/47.

DR P-PSDB; AAM41760.

XX Novel nucleic acids and polypeptides, useful for treating disorders such

PT as central nervous system injuries.

XX Claim 1; SEQ ID NO 4905; 10078pp; English.

XX The invention relates to human nucleic acids (AAI57798-AAI61369) and the

CC encoded polypeptides (AAM38642-AAM42213) with nootropic,

CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Actin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders. Note: The sequence data for this patent did not form  
CC part of the printed specification  
XX  
XX  
SQ Sequence 1401 BP; 473 A; 286 C; 320 G; 322 T; 0 U; 0 Other;

Query Match 83.2%; Score 15.8; DB 4; Length 1401;  
Best Local Similarity 89.5%; Pred. No. 1.6e+03;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCCTTTG 19  
||||| ||||| |||||  
Db 829 CTTTATTTCATAGCCTTTG 811

RESULT 73  
ADB83252/c  
ID ADB83252 standard; cDNA; 1599 BP.  
XX  
AC ADB83252;  
XX  
XX  
DT 04-DEC-2003 (first entry)  
XX  
DE Human cDNA sequence useful for the treatment of cancer (SeqID 1465).  
XX  
XX human; prostate; cancer; cytostatic; gene therapy; vaccine;  
KW immune response; gene; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO2003050236-A2.  
XX  
XX 19-JUN-2003.  
XX  
XX 04-SEP-2002; 2002WO-US028214.  
XX  
XX 07-DEC-2001; 2001US-00012697.  
XX  
XX (CHIR ) CHIRON CORP.  
PA (HYSE-) HYSEQ INC.PA

Escobedo J, Garcia PD, Kassam A, Lamson G, Drmanac R;  
PI Crkvenjakov R, Dickson M, Drmanac S, Labat I, Leehkowitz D, Kita D;  
PI Garcia V, Jones LW, Stache-Crain B, Scott EM;  
XX  
XX WPI; 2003-513972/48.  
XX  
XX New polynucleotides derived from human prostate, useful for modulating  
PT immune response to prevent or treat cancer.  
PT  
XX  
XX Claim 1; SEQ ID NO 1465; 189pp; English.

XX This invention relates to novel isolated polynucleotides of human origin,  
CC particularly isolated from the human prostate. Specifically, it refers to  
CC the diagnostics and therapeutics comprising these novel human  
CC polynucleotides, and includes the derived probes, antisense  
CC oligonucleotides and antibodies thereof. The identification of these  
CC human prostate genes that can inhibit tumour growth is useful for  
CC understanding the progression and nature of complex diseases such as  
CC cancer, and hence they are important in the drug discovery process. The  
CC present invention describes these polynucleotides and encoded  
CC polypeptides as exhibiting cytostatic activity, and through gene therapy  
CC and/ or vaccines they can be used to modulate the immune response for the

CC prevention or treatment of cancers, particularly of the prostate, but  
CC also for breast, lung and colon cancer. This polynucleotide sequence is a  
CC human cDNA sequence useful for the treatment of cancer, used in an  
CC exemplification of the invention. NOTE: These sequences are not given in  
CC the specification but are provided on the WIPO website.

XX  
SQ Sequence 1599 BP; 528 A; 327 C; 378 G; 366 T; 0 U; 0 Other;  
Query Match 83.2%; Score 15.8; DB 9; Length 1599;  
Best Local Similarity 89.5%; Pred. No. 1.6e+03;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCCTTTG 19  
||||| ||||| |||||  
Db 456 CTTTATTTCATAGCCTTTG 438

RESULT 74  
AAS64753/c  
ID AAS64753 standard; cDNA; 1734 BP.  
XX  
XX AAS64753;  
XX  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #557.  
XX  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200175067-A2.  
XX  
XX 11-OCT-2001.  
XX  
XX 30-MAR-2001; 2001WO-US008631.  
XX  
XX 31-MAR-2000; 2000US-00540217.  
PR 23-AUG-2000; 2000US-00649167.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Drmanac RT, Liu C, Tang YT;  
PI WPI; 2001-639362/73.  
DR P-PSDB; ABG00566.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX  
XX Claim 1; SEQ ID NO 557; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic  
CC coding sequences of the invention. Note: The sequence data for this

CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 1734 BP; 507 A; 419 C; 458 G; 350 T; 0 U; 0 Other;  
 Best Local Similarity 83.2%; Score 15.8; DB 5; Length 1734;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 CTTTACTTCATAGCTTTG 19  
 ||||| ||||| |||||  
 Db 1272 CTTTATTTCATAGCCTTTG 1254

RESULT 75  
 ADN38895/c  
 ID ADN38895 standard; cDNA; 1758 BP.

XX AC ADN38895;

XX DT 17-JUN-2004 (first entry)

XX DE Cancer/angiogenesis/fibrosis-related nucleic acid, SEQ ID NO:213.

XX KW Human; differential expression; cancer; angiogenic disorder;  
 KW fibrotic disorder; psoriasis; ischaemia; heart disease; atherosclerosis;  
 KW inflammatory disease; autoimmune disease;  
 KW retinal neovascularisation syndrome; scarring; uterine fibroid;  
 KW detection; diagnosis; prognosis; drug screening; drug targeting;  
 KW wound healing; contraception; cytostatic; cardiant; immunomodulatory;  
 KW vulnary; gene therapy; vaccine; gene; ss.

XX OS Homo sapiens.

XX PN WO2003042661-A2.

XX PD 22-MAY-2003.

XX PF 13-NOV-2002; 2002WO-US036810.

XX PR 13-NOV-2001; 2001US-0350666P.

XX PR 21-NOV-2001; 2001US-0332464P.

XX PR 29-NOV-2001; 2001US-0334393P.

XX PR 03-DEC-2001; 2001US-0335394P.

XX PR 14-DEC-2001; 2001US-0340376P.

XX PR 08-JAN-2002; 2002US-0347211P.

XX PR 10-JAN-2002; 2002US-0347349P.

XX PR 08-FEB-2002; 2002US-0355250P.

XX PR 13-FEB-2002; 2002US-0356714P.

XX PR 20-FEB-2002; 2002US-0359077P.

XX PR 29-MAR-2002; 2002US-0368809P.

XX PR 04-APR-2002; 2002US-0370110P.

XX PR 12-APR-2002; 2002US-0372246P.

XX PR 05-JUN-2002; 2002US-0386614P.

XX PR 16-JUL-2002; 2002US-0396839P.

XX PR 22-JUL-2002; 2002US-0397775P.

XX PR 22-JUL-2002; 2002US-0397845P.

XX PR 09-SEP-2002; 2002US-0409450P.

XX PA (EOSB-) EOS BIOTECHNOLOGY INC.

XX PI Afar D, Aziz N, Ginsburg WM, Gish KC, Glynn R, Hevezi PA;

XX PI Mack DH, Murray R, Watson SR, Wilson KE, Zlotnik A;

XX DR WPI; 2003-468649/44.

XX DR P-PSDB; ADN38896.

XX FT Determining the presence or absence of a pathological cell in a patient,  
 PT useful for diagnosing, prognosing or treating cancer, comprises detecting  
 PT a nucleic acid in a biological sample.

XX PS Claim 8; SEQ ID NO 213; 1385pp; English.

XX CC The invention relates to nucleic acids and proteins (ADN38683-ADN40064)  
 CC whose expression is upregulated or downregulated in specific cancers or  
 CC other diseases such as angiogenic or fibrotic disorders, and to methods  
 CC of determining the presence or absence of a pathological cell in a  
 CC patient by detecting a nucleic acid at least 80% identical to those of  
 CC the invention or by detecting a polypeptide of the invention. The  
 CC invention also relates to expression vectors and host cells comprising a  
 CC nucleic acid of the invention; antibodies which specifically bind a  
 CC polypeptide of the invention; use of such antibodies for drug targeting;  
 CC and methods of screening for modulators of activity or expression of the  
 CC polypeptides and nucleic acids. The nucleic acids, polypeptides,  
 CC antibodies and methods are useful for diagnosing, prognosing and treating  
 CC cancer and other conditions such as psoriasis, ischaemia, heart disease,  
 CC atherosclerosis, inflammatory diseases, autoimmune diseases, retinal  
 CC neovascularisation syndromes, scarring and uterine fibroids. They may  
 CC also be useful in wound healing and in contraception. The present  
 CC sequence represents a nucleic acid sequence of the invention.  
 XX SQ Sequence 1758 BP; 554 A; 372 C; 412 G; 420 T; 0 U; 0 Other;

Query Match 83.2%; Score 15.8; DB 11; Length 1758;  
 Best Local Similarity 89.5%; Pred. No. 1.6e+03;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 CTTTACTTCATAGCTTTG 19  
 ||||| ||||| |||||  
 Db 615 CTTTATTTCATAGCCTTTG 597

RESULT 76  
 ACC43802/c  
 ID ACC43802 standard; DNA; 1868 BP.

XX AC ACC43802;

XX DT 11-AUG-2003 (first entry)

XX DE Nucleotide sequence of plant regulatory polynucleotide.

XX KW Plant; regulatory polynucleotide; endothelium; albumen; seed; seed testa;  
 KW flavonoid; tannin; cellulose; maize; rape seed; colour; nutrition; ss.

XX OS Arabidopsis thaliana.

XX PN FR2828210-A1.

XX PD 07-FEB-2003.

XX PF 01-AUG-2001; 2001FR-00010365.

XX PR 01-AUG-2001; 2001FR-00010365.

XX PA (INRG ) INRA INST NAT RECH AGRONOMIQUE.  
 XX (CNRS ) CNRS CENT NAT RECH SCI.

XX PI Lepintec L, Debeaujon I, Devic M, Caboche M;

XX DR WPI; 2003-335321/32.

XX PT New regulatory nucleic acid, useful for preparing transgenic plants with  
 PT altered seed testa, provides specific expression of attached genes in  
 XX seed endothelium and albumen.

XX PS Disclosure; Page 80; 85pp; French.

XX CC The present sequence represents a plant regulatory polynucleotide, which  
 CC was used to construct plasmid pBan2. The present sequence is a fragment  
 CC of the sequence given in ACC43789, which comprises regulatory sequences  
 CC that allow expression of a polynucleotide specifically in the endothelium  
 CC and/or albumen of plant seeds. The regulatory polynucleotide, and derived  
 CC expression cassettes or vectors, are used to produce transgenic plants in  
 CC which development of seed testa is altered, especially for reducing the

CC size or number of seeds, particularly complete absence of mature, fertile  
CC seeds, or modulating contents of flavonoids, tannins and cellulose  
CC (especially to produce maize and rape seeds with less intense yellow  
CC colour, or to alter nutritional properties)

XX Sequence 1868 BP; 592 A; 304 C; 402 G; 570 T; 0 U; 0 Other;

SQ Query Match 83.2%; Score 15.8; DB 8; Length 1868;

Best Local Similarity 89.5%; Pred. No. 1.6e+03;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCTTTG 19

|||||  
Db 532 CTTCACTCCATAGTCTTTG 514

RESULT 77

ADJ41386/C

ID ADJ41386 standard; cDNA; 2000 BP.

XX AC

ADJ41386;

XX 06-MAY-2004 (first entry)

XX Plant cDNA #2386.

DE

XX Plant; gene; ss; transcripition; plant genome augmentation; cereal;  
KW soybean; alfalfa; sunflower; canola; cotton; peanut; tobacco; sugar beet;  
KW maize; barley; sorghum; rice; wheat; crop plant; insecticide resistance;  
KW stress tolerance; salt tolerance; cold tolerance; drought tolerance;  
KW plant nutrition; apical dominance; dwarfism; early flowering; antiviral;  
KW antifungal.

XX Eukaryota.

XX OS

XX US2004016025-A1.

XX 23-JAN-2004.

XX 26-SEP-2002; 2002US-00260238.

XX 26-SEP-2001; 2001US-0325277P.

XX 26-SEP-2001; 2001US-0325448P.

XX 04-APR-2002; 2002US-0370620P.

XX (BUDW/) BUDWORTH P.

XX (MOUG/) MOUGHAMER T.

XX (BRIG/) BRIGGS S P.

XX (COOP/) COOPER B.

XX (GLAZ/) GLAZEBROOK J.

XX (GOFF/) GOFF S A.

XX (KATA/) KATAGIRI F.

XX (KREP/) KREPS J.

XX (PROV/) PROVART N.

XX (RICK/) RICHE D.

XX (ZHUT/) ZHU T.

XX Budworth P, Moughamer T, Briggs SP, Cooper B, Glazebrook J;

XX Goff SA, Katagiri F, Kreps J, Provart N, Ricke D, Zhu T;

XX WPI; 2004-190374/18.

XX New rice promoter, useful for manipulating crop plants to alter or

XX improve phenotypic characteristics, e.g. produce large quantities of oil

XX or proteins, resistance to insecticides, virus or fungi, stress tolerance

XX or high nutritional value.

XX Claim 1; SEQ ID NO 2386; 230pp; English.

XX The invention relates to plant nucleotide sequences that direct seed-,

XX leaf- and/or stem-, panicle-, root- or pollen-specific or -preferential

XX or constitutive transcription of an operatively linked nucleic acid

XX segment. The invention also relates to a method for augmenting a plant

CC genome and a method of identifying a gene, where its expression is  
CC altered in the seed, leaf, stem, panicle, pollen, root or is constitutive  
CC in a plant cell. The plant is a cereal, e.g. soybean, alfalfa, sunflower,  
CC canola, cotton, peanut, tobacco or sugar beet, preferably maize, barley,  
CC sorghum, rice or wheat. The polynucleotides and the polypeptides they  
CC encode are useful for manipulating crop plants to alter or improve  
CC phenotypic characteristics, to produce large quantities of oil or  
CC proteins, to incur resistance to insecticides, viruses or fungi, and to  
CC incur stress tolerance (e.g. salt, cold or drought) to ensure the plants  
CC have a high nutritional value with reduced apical dominance or dwarfism,  
CC early flowering or altered metabolic pathways. This sequence represents a  
CC plant nucleic acid of the invention. Note: The sequence data for this  
CC patent did not form part of the printed specification but was obtained in  
CC electronic format directly from USPTO at [seqdata.uspto.gov/sequence.html](http://seqdata.uspto.gov/sequence.html).

XX SQ Sequence 2000 BP; 542 A; 542 C; 499 G; 417 T; 0 U; 0 Other;

Query Match 83.2%; Score 15.8; DB 12; Length 2000;

Best Local Similarity 89.5%; Pred. No. 1.7e+03;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCTTTG 19

|||||  
Db 675 CTTTGCTTCATAGTCTTCG 657

RESULT 78

ABL60217/C

ID ABL60217 standard; cDNA; 2021 BP.

XX AC

ABL60217;

XX 29-JUL-2002 (first entry)

XX Human encoding zinc finger protein 59 cDNA SEQ ID NO 1.

XX Human; zinc finger protein 59; disease; gene; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 191..1789

XX /\*tag= a

XX /product= "zinc finger protein 59"

XX CN1293203-A.

XX 02-MAY-2001.

XX 18-OCT-1999; 99CN-00116988.

XX 18-OCT-1999; 99CN-00116988.

XX (SHAN-) SHANGHAI BORONG GENE DEV CO LTD.

XX Mao Y, Xie Y;

XX WPI; 2001-418901/45.

XX P-PSDB; ABB77494.

XX Polypeptide-human zinc finger protein 59 and polynucleotide for coding

XX this polypeptide.

XX Claim 6; Page 16-17 (Disclosure); 25pp; Chinese.

XX The invention relates to human zinc finger protein 59 (ABB77494), the

XX polynucleotide (ABL60217) encoding it, the process for preparing it, its

XX application in treating several diseases, the antagonist and medical

XX action and coding the new applications of the polynucleotide are

XX disclosed

XX Sequence 2021 BP; 680 A; 398 C; 478 G; 465 T; 0 U; 0 Other;

```
* Query Match      83.2%; Score 15.8; DB 4; Length 2021;
Best Local Similarity 89.5%; Pred. No. 1.7e+03;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19
    ||||| ||||| |||||
Db 646 CTTTATTTCATAGCCTTTG 628

RESULT 79
AAH16484/C
ID AAH16484 standard; cDNA; 2063 BP.
XX
AC AAH16484;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:15507.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
FN EP1074617-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-00116126.
XX
XX 29-JUL-1999; 99JP-00248036.
XX
XX 27-AUG-1999; 99JP-00300253.
XX
XX 11-JAN-2000; 2000JP-00118776.
XX
XX 02-MAY-2000; 2000JP-00183767.
XX
XX 09-JUN-2000; 2000JP-00241899.
XX
PA (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX
XX Iehii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-
XX length cDNAs defined in the specification, and for the detection and/or
XX diagnosis of the abnormality of the proteins encoded by the full-length
XX cDNAs.
XX
XX Claim 8; SEQ ID NO 15507; 2537pp + Sequence Listing; English.
XX
XX The present invention describes primer sets for synthesising 5602 full-
XX length cDNAs defined in the specification. Where a primer set comprises:
XX (a) an oligo-dT primer and an oligonucleotide complementary to the
XX complementary strand of a polynucleotide which comprises one of the 5602
XX nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in the
XX specification. The primer sets can be used in antisense therapy and in
XX gene therapy. The primers are useful for synthesising polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893
XX represent human amino acid sequences; and AAH13629 to AAH13632 represent
XX oligonucleotides, all of which are used in the exemplification of the
XX present invention
XX
XX Sequence 2063 BP; 691 A; 406 C; 489 G; 477 T; 0 U; 0 Other;

Query Match      83.2%; Score 15.8; DB 4; Length 2021;
Best Local Similarity 89.5%; Pred. No. 1.7e+03;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19
    ||||| ||||| |||||
Db 646 CTTTATTTCATAGCCTTTG 628

RESULT 79
AAH16484/C
ID AAH16484 standard; cDNA; 2063 BP.
XX
AC AAH16484;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:15507.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
FN EP1074617-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-00116126.
XX
XX 29-JUL-1999; 99JP-00248036.
XX
XX 27-AUG-1999; 99JP-00300253.
XX
XX 11-JAN-2000; 2000JP-00118776.
XX
XX 02-MAY-2000; 2000JP-00183767.
XX
XX 09-JUN-2000; 2000JP-00241899.
XX
PA (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX
XX Iehii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-
XX length cDNAs defined in the specification, and for the detection and/or
XX diagnosis of the abnormality of the proteins encoded by the full-length
XX cDNAs.
XX
XX Claim 8; SEQ ID NO 15507; 2537pp + Sequence Listing; English.
XX
XX The present invention describes primer sets for synthesising 5602 full-
XX length cDNAs defined in the specification. Where a primer set comprises:
XX (a) an oligo-dT primer and an oligonucleotide complementary to the
XX complementary strand of a polynucleotide which comprises one of the 5602
XX nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in the
XX specification. The primer sets can be used in antisense therapy and in
XX gene therapy. The primers are useful for synthesising polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893
XX represent human amino acid sequences; and AAH13629 to AAH13632 represent
XX oligonucleotides, all of which are used in the exemplification of the
XX present invention
XX
XX Sequence 2063 BP; 691 A; 406 C; 489 G; 477 T; 0 U; 0 Other;

Query Match      83.2%; Score 15.8; DB 4; Length 2063;
Best Local Similarity 89.5%; Pred. No. 1.7e+03;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19
    ||||| ||||| |||||
Db 684 CTTTATTTCATAGCCTTTG 666

RESULT 80
ABL20545
ID ABL20545 standard; DNA; 2094 BP.
XX
AC ABL20545;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 13108.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
XX
XX 11-JUL-2000; 2000US-00614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions.
XX
XX Claim 1; SEQ ID NO 13108; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-
XX ABBS72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 2094 BP; 478 A; 569 C; 546 G; 501 T; 0 U; 0 Other;

Query Match      83.2%; Score 15.8; DB 4; Length 2094;
Best Local Similarity 89.5%; Pred. No. 1.7e+03;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19
    ||||| ||||| |||||
Db 784 CTCTACTTCATAGTCTTTG 802

RESULT 81
ABL05491
ID ABL05491 standard; cDNA; 2165 BP.
XX
XX
AC ABL05491;
```

XX DT 26-MAR-2002 (first entry)  
XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 10955.  
XX DE Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX OS Drosophila melanogaster.  
XX PN WO200171042-A2.  
XX PD 27-SEP-2001.  
XX PF 23-MAR-2001; 2001WO-US009231.  
XX PR 23-MAR-2000; 2000US-0191637P.  
XX PR 11-JUL-2000; 2000US-00614150.  
XX PA (PEKE ) PE CORP NY.  
XX PI Venter JC, Adams M, Li PWD, Myers EW;  
XX WPI; 2001-656860/75.  
DR P-PSDB; ABB61388.  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.  
XX Claim 1; SEQ ID NO 10955; 21pp + Sequence Listing; English.  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pt\_sequences  
XX SQ Sequence 2165 BP; 501 A; 590 C; 556 G; 518 T; 0 U; 0 Other;  
Query Match 83.2%; Score 15.8; DB 4; Length 2165;  
Best Local Similarity 89.5%; Pred. No. 1.7e+03;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CTTTACTTCATAGTCTTTG 19  
Db 855 CTCTACTTCATGCTTTG 873  
RESULT 82  
ID ABQ72580/c  
AC ABQ72580 standard; cDNA; 2175 BP.  
XX AC ABQ72580;  
XX DT 03-SEP-2002 (first entry)  
XX Human MDDT encoding cDNA SEQ ID NO 132.  
XX Human; MDDT; disease detection and treatment molecule polynucleotide;  
KW proliferative disorder; hepatitis; psoriasis; cancer; AIDS;  
KW autoimmune disorder; inflammatory disorder; allergy; multiple sclerosis;  
KW rheumatoid arthritis; transgenic; gene therapy; antiatherosclerotic;  
KW hepatotropic; antiinflammatory; antipsoriatic; cytostatic; anti-HIV;  
KW antiallergic; antianaemic; antiasthmatic; antiatherosclerotic; antigout;  
KW neuroprotective; antirheumatic; antiarthritic; gene; ss.  
XX Homo sapiens.

XX WO200240715-A2.  
XX 23-MAY-2002.  
XX 06-SEP-2001; 2001WO-US027628.  
XX 05-SEP-2000; 2000US-0229747P.  
XX 05-SEP-2000; 2000US-0229748P.  
XX 05-SEP-2000; 2000US-0229749P.  
XX 05-SEP-2000; 2000US-0229750P.  
XX 05-SEP-2000; 2000US-0229751P.  
XX 05-SEP-2000; 2000US-0230583P.  
XX 05-SEP-2000; 2000US-0230585P.  
XX 05-SEP-2000; 2000US-0230514P.  
XX 05-SEP-2000; 2000US-0230515P.  
XX 05-SEP-2000; 2000US-0230517P.  
XX 05-SEP-2000; 2000US-0230518P.  
XX 05-SEP-2000; 2000US-0230519P.  
XX 05-SEP-2000; 2000US-0230595P.  
XX 05-SEP-2000; 2000US-0230597P.  
XX 05-SEP-2000; 2000US-0230598P.  
XX 05-SEP-2000; 2000US-0230599P.  
XX 05-SEP-2000; 2000US-0230610P.  
XX 05-SEP-2000; 2000US-0230865P.  
XX 05-SEP-2000; 2000US-0230988P.  
XX 07-SEP-2000; 2000US-0230989P.  
XX 07-SEP-2000; 2000US-0231163P.  
XX 07-SEP-2000; 2000US-0231167P.  
XX (INCY-) INCYTE GENOMICS INC.  
XX Jackson S, Lincoln SE, Altus CM, Dufour GE, Chalup MS;  
XX Hillman JL, Jones AL, Yu JY, Wright RJ, Gietzen D, Liu TF, Yap PE;  
XX Dahl CR, Momiyama MG, Bradley DL, Rohatgi SD, Harris B;  
XX Roseberry AM, Gerstin EH, Peralta CH, David MH, Panzer SR, Flores V;  
XX Daifo A, Marwaha R, Chen AJ, Chang SC, Au AP, Inman RR;  
XX WPI; 2002-527544/56.  
XX P-PSDB; ABP51363.  
XX Novel human disease detection and treatment polypeptide, useful in  
PT diagnosis, prevention or treatment of cell proliferative disorders e.g.  
PT arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder e.g.  
PT AIDS.  
XX Claim 1; Page 381; 618pp; English.  
XX The invention relates to an isolated human disease detection and  
CC treatment (MDDT) polypeptide (I) selected from a polypeptide having a  
CC sequence selected from 254 sequences (ABP51231-ABP51484) given in the  
CC specification, a naturally occurring polypeptide comprising a sequence  
CC having at least 90% identity to (I) or a biologically active or  
CC immunogenic fragment of (I). (I) is useful for screening a compound for  
CC effectiveness as an agonist or antagonist, for screening a compound that  
CC specifically binds (I) or modulates the activity of (I), and for  
CC preparing a polyclonal or monoclonal antibody by hybridoma technology.  
CC Nucleic acids (II) (ABQ72449-ABQ72700) encoding (I) are useful for  
CC screening a compound for effectiveness in altering expression of a target  
CC polynucleotide comprising. Oligonucleotides and antibodies are useful for  
CC detecting MDDT in a sample or for assessing toxicity of a test compound,  
CC in a diagnostic test for a condition or a disease associated with the  
CC expression of MDDT in a biological sample, for detecting (I) in a sample,  
CC and for purifying (I) from a sample. A composition comprising (I), an  
CC agonist or antagonist is useful for treating a disease or condition  
CC associated with decreased or increased expression of functional MDDT. (I)  
CC or (II) are useful for diagnosing, treating or preventing disorders  
CC associated with aberrant expression of MDDT, where the disorders are  
CC selected from a cell proliferative disorder such as arteriosclerosis,  
CC cirrhosis, hepatitis, psoriasis, and cancer and an  
CC autoimmune/inflammatory disorder such as AIDS, Addison's disease,  
CC allergy, anaemia, asthma, atherosclerosis, gout, multiple sclerosis or



CC rheumatoid arthritis. (II) are useful for creating knockin humanised  
 CC animals or transgenic animals to model human diseases, in somatic or  
 CC germline gene therapy, to generate a transcript image of a tissue or cell  
 CC type, for detecting differences in the chromosomal location due to  
 CC translocation or inversion among normal, carrier or affected individuals  
 CC and as hybridisation probes for mapping naturally occurring genomic  
 CC sequences

SQ Sequence 2175 BP; 726 A; 415 C; 526 G; 508 T; 0 U; 0 Other;

Query Match 83.2%; Score 15.8; DB 6; Length 2175;  
 Best Local Similarity 89.5%; Pred. No. 1.7e+03;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGCTTTG 19  
 ||||| ||||| ||||| |||||  
 Db 677 CTTTATTCATAGCCTTTG 659

RESULT 83

ID AD226902/c  
 ID AD226902 standard; DNA; 2208 BP.

XX AC AD226902;

XX DT 30-JUN-2005 (first entry)

XX DE Adeno-associated virus DNA SEQ ID NO 52.

XX KW rheumatoid arthritis; multiple sclerosis; sarcoidosis; diabetes;  
 KW scleroderma; psoriasis; vasculitis; Crohn's disease; hemophilia; HIV;  
 KW bacterial infection; cancer; ulcerative colitis; antirheumatic;  
 KW antiarthritic; neuroprotective; antiinflammatory; antidiabetic;  
 KW antipsoriatic; vasotropic; gastrointestinal-gen.; hemostatic; anti-HIV;  
 KW virucide; antibacterial; cytostatic; antiulcer; dermatological; ds.

XX OS Adeno-associated virus.

XX PN WO2005033321-A2.

XX PD 14-APR-2005.

XX PF 30-SEP-2004; 2004WO-US028817.

XX PR 30-SEP-2003; 2003US-0508226P.

XX PR 29-APR-2004; 2004US-0566546P.

XX PA (UYPE-) UNIV PENNSYLVANIA.

XX PI Wilson JM, Gao G, Alvira MR, Vandenberghe LH;

XX DR WPI; 2005-285437/29.

XX CC New adeno-associated virus (AAV) clade comprising at least three AAV  
 PT members, useful for preventing and/or treating arthritis, multiple  
 PT sclerosis, diabetes, scleroderma, psoriasis, hemophilia, HIV, bacterial  
 PT infection and cancer.

XX PS Claim 46; SEQ ID NO 52; 569pp; English.

XX CC The invention relates to an adeno-associated virus (AAV) clade comprising  
 CC at least three AAV members, where each member of the AAV clade is  
 CC phylogenetically related as determined using a Neighbor-Joining heuristic  
 CC by a bootstrap value of at least 75 % per 1000 isolates and a Poisson  
 CC correction distance measurement of no more than 0.05. The methods and  
 CC compositions of the present invention are useful for the prevention  
 CC and/or treatment of rheumatoid arthritis, multiple sclerosis,  
 CC sarcoidosis, diabetes, scleroderma, psoriasis, vasculitis, Crohn's  
 CC disease, hemophilia, HIV, bacterial infection, cancer and ulcerative  
 CC colitis. The present sequence represents an adeno-associated virus DNA.

XX SQ Sequence 2208 BP; 599 A; 629 C; 526 G; 454 T; 0 U; 0 Other;

Query Match 83.2%; Score 15.8; DB 14; Length 2208;  
 Best Local Similarity 89.5%; Pred. No. 1.7e+03;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGCTTTG 19  
 ||||| ||||| ||||| |||||  
 Db 1596 CTTTCTTCATCGCTTTG 1578

RESULT 84

ID AD226918/c  
 ID AD226918 standard; DNA; 2208 BP.

XX AC AD226918;

XX DT 30-JUN-2005 (first entry)

XX DE Adeno-associated virus DNA SEQ ID NO 68.

XX KW rheumatoid arthritis; multiple sclerosis; sarcoidosis; diabetes;  
 KW scleroderma; psoriasis; vasculitis; Crohn's disease; hemophilia; HIV;  
 KW bacterial infection; cancer; ulcerative colitis; antirheumatic;  
 KW antiarthritic; neuroprotective; antiinflammatory; antidiabetic;  
 KW antipsoriatic; vasotropic; gastrointestinal-gen.; hemostatic; anti-HIV;  
 KW virucide; antibacterial; cytostatic; antiulcer; dermatological; ds.

XX OS Adeno-associated virus.

XX PN WO2005033321-A2.

XX PD 14-APR-2005.

XX PF 30-SEP-2004; 2004WO-US028817.

XX PR 30-SEP-2003; 2003US-0508226P.

XX PR 29-APR-2004; 2004US-0566546P.

XX PA (UYPE-) UNIV PENNSYLVANIA.

XX PI Wilson JM, Gao G, Alvira MR, Vandenberghe LH;

XX DR WPI; 2005-285437/29.

XX CC New adeno-associated virus (AAV) clade comprising at least three AAV  
 PT members, useful for preventing and/or treating arthritis, multiple  
 PT sclerosis, diabetes, scleroderma, psoriasis, hemophilia, HIV, bacterial  
 PT infection and cancer.

XX PS Claim 15; SEQ ID NO 68; 569pp; English.

XX CC The invention relates to an adeno-associated virus (AAV) clade comprising  
 CC at least three AAV members, where each member of the AAV clade is  
 CC phylogenetically related as determined using a Neighbor-Joining heuristic  
 CC by a bootstrap value of at least 75 % per 1000 isolates and a Poisson  
 CC correction distance measurement of no more than 0.05. The methods and  
 CC compositions of the present invention are useful for the prevention  
 CC and/or treatment of rheumatoid arthritis, multiple sclerosis,  
 CC sarcoidosis, diabetes, scleroderma, psoriasis, vasculitis, Crohn's  
 CC disease, hemophilia, HIV, bacterial infection, cancer and ulcerative  
 CC colitis. The present sequence represents an adeno-associated virus DNA.

SQ Sequence 2208 BP; 592 A; 641 C; 544 G; 431 T; 0 U; 0 Other;

Query Match 83.2%; Score 15.8; DB 14; Length 2208;  
 Best Local Similarity 89.5%; Pred. No. 1.7e+03;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGCTTTG 19  
 ||||| ||||| ||||| |||||  
 Db 1596 CTTTCTTCATCGCTTTG 1578

RESULT 85

```

AD226920/c
ID AD226920 standard; DNA; 2208 BP.
XX
AC AD226920;
XX
XX 30-JUN-2005 (first entry)
DT
XX
DE Adeno-associated virus DNA SEQ ID NO 70.
DE
XX
KW rheumatoid arthritis; multiple sclerosis; sarcoidosis; diabetes;
KW scleroderma; psoriasis; vasculitis; Crohn's disease; hemophilia; HIV;
KW bacterial infection; cancer; ulcerative colitis; antirheumatic;
KW antiarthritic; neuroprotective; antinflammatory; antidiabetic;
KW antipsoriatic; vasotropic; gastrointestinal-gen.; hemostatic; anti-HIV;
KW virucide; antibacterial; cytostatic; antiulcer; dermatological; ds.
XX
OS Adeno-associated virus.
XX
PN WO2005033321-A2.
XX
XX 14-APR-2005.
XX
XX 30-SEP-2004; 2004WO-US028817.
XX
XX 30-SEP-2003; 2003US-0508226P.
XX
XX 29-APR-2004; 2004US-0566546P.
XX
XX (UYPE-) UNIV PENNSYLVANIA.
XX
XX Wilson JM, Gao G, Alvira MR, Vandenberghe LH;
XX WPI; 2005-285437/29.
XX
XX New adeno-associated virus (AAV) clade comprising at least three AAV
XX members, useful for preventing and/or treating arthritis, multiple
XX sclerosis, diabetes, scleroderma, psoriasis, hemophilia, HIV, bacterial
XX infection and cancer.
XX
XX Claim 50; SEQ ID NO 70; 569pp; English.
XX
XX The invention relates to an adeno-associated virus (AAV) clade comprising
XX at least three AAV members, where each member of the AAV clade is
XX phylogenetically related as determined using a Neighbor-Joining heuristic
XX by a bootstrap value of at least 75 % per 1000 isolates and a Poisson
XX correction distance measurement of no more than 0.05. The methods and
XX compositions of the present invention are useful for the prevention
XX and/or treatment of rheumatoid arthritis, multiple sclerosis,
XX sarcoidosis, diabetes, scleroderma, psoriasis, vasculitis, Crohn's
XX disease, hemophilia, HIV, bacterial infection, cancer and ulcerative
XX colitis. The present sequence represents an adeno-associated virus DNA.
XX
XX Sequence 2208 BP; 587 A; 640 C; 548 G; 433 T; 0 U; 0 Other;
XX
Query Match 83.2%; Score 15.8; DB 14; Length 2208;
Best Local Similarity 89.5%; Pred. No. 1.7e+03;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
QY 1 CTTTACTTCATAGCTTTTG 19
DB 1596 CTTTCTTCATCGCTTTTG 1578

RESULT 86
AD226907/c
ID AD226907 standard; DNA; 2208 BP.
XX
AC AD226907;
XX
XX 30-JUN-2005 (first entry)
DT
XX
DE Adeno-associated virus DNA SEQ ID NO 57.
DE
XX
KW rheumatoid arthritis; multiple sclerosis; sarcoidosis; diabetes;
KW scleroderma; psoriasis; vasculitis; Crohn's disease; hemophilia; HIV;
KW bacterial infection; cancer; ulcerative colitis; antirheumatic;
KW antiarthritic; neuroprotective; antinflammatory; antidiabetic;
KW antipsoriatic; vasotropic; gastrointestinal-gen.; hemostatic; anti-HIV;
KW virucide; antibacterial; cytostatic; antiulcer; dermatological; ds.
XX
OS Adeno-associated virus.
XX
PN WO2005033321-A2.
XX
XX 14-APR-2005.
XX
XX 30-SEP-2004; 2004WO-US028817.
XX
XX 30-SEP-2003; 2003US-0508226P.
XX
XX 29-APR-2004; 2004US-0566546P.
XX
XX (UYPE-) UNIV PENNSYLVANIA.
XX
XX Wilson JM, Gao G, Alvira MR, Vandenberghe LH;
XX WPI; 2005-285437/29.
XX
XX New adeno-associated virus (AAV) clade comprising at least three AAV
XX members, useful for preventing and/or treating arthritis, multiple
XX sclerosis, diabetes, scleroderma, psoriasis, hemophilia, HIV, bacterial
XX infection and cancer.
XX
XX Disclosure; SEQ ID NO 70; 569pp; English.
XX
XX The invention relates to an adeno-associated virus (AAV) clade comprising
XX at least three AAV members, where each member of the AAV clade is
XX phylogenetically related as determined using a Neighbor-Joining heuristic
XX by a bootstrap value of at least 75 % per 1000 isolates and a Poisson
XX correction distance measurement of no more than 0.05. The methods and
XX compositions of the present invention are useful for the prevention
XX and/or treatment of rheumatoid arthritis, multiple sclerosis,
XX sarcoidosis, diabetes, scleroderma, psoriasis, vasculitis, Crohn's
XX disease, hemophilia, HIV, bacterial infection, cancer and ulcerative
XX colitis. The present sequence represents an adeno-associated virus DNA.
XX
XX Sequence 2208 BP; 587 A; 640 C; 548 G; 433 T; 0 U; 0 Other;
XX
Query Match 83.2%; Score 15.8; DB 14; Length 2208;
Best Local Similarity 89.5%; Pred. No. 1.7e+03;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
QY 1 CTTTACTTCATAGCTTTTG 19
DB 1596 CTTTCTTCATCGCTTTTG 1578

RESULT 87
AD226903/c
ID AD226903 standard; DNA; 2208 BP.
XX
AC AD226903;
XX
XX 30-JUN-2005 (first entry)
DT
XX
DE Adeno-associated virus DNA SEQ ID NO 53.
DE
XX
KW rheumatoid arthritis; multiple sclerosis; sarcoidosis; diabetes;
KW scleroderma; psoriasis; vasculitis; Crohn's disease; hemophilia; HIV;
KW bacterial infection; cancer; ulcerative colitis; antirheumatic;
KW antiarthritic; neuroprotective; antinflammatory; antidiabetic;
KW antipsoriatic; vasotropic; gastrointestinal-gen.; hemostatic; anti-HIV;
KW virucide; antibacterial; cytostatic; antiulcer; dermatological; ds.
XX
OS Adeno-associated virus.
XX
PN WO2005033321-A2.
XX
XX 14-APR-2005.
XX
XX 30-SEP-2004; 2004WO-US028817.
XX
XX 30-SEP-2003; 2003US-0508226P.
XX
XX 29-APR-2004; 2004US-0566546P.
XX
XX (UYPE-) UNIV PENNSYLVANIA.
XX
XX Wilson JM, Gao G, Alvira MR, Vandenberghe LH;
XX WPI; 2005-285437/29.
XX
XX New adeno-associated virus (AAV) clade comprising at least three AAV
XX members, useful for preventing and/or treating arthritis, multiple
XX sclerosis, diabetes, scleroderma, psoriasis, hemophilia, HIV, bacterial
XX infection and cancer.
XX
XX Claim 50; SEQ ID NO 57; 569pp; English.
XX
XX The invention relates to an adeno-associated virus (AAV) clade comprising
XX at least three AAV members, where each member of the AAV clade is
XX phylogenetically related as determined using a Neighbor-Joining heuristic
XX by a bootstrap value of at least 75 % per 1000 isolates and a Poisson
XX correction distance measurement of no more than 0.05. The methods and
XX compositions of the present invention are useful for the prevention
XX and/or treatment of rheumatoid arthritis, multiple sclerosis,
XX sarcoidosis, diabetes, scleroderma, psoriasis, vasculitis, Crohn's
XX disease, hemophilia, HIV, bacterial infection, cancer and ulcerative
XX colitis. The present sequence represents an adeno-associated virus DNA.
XX
XX Sequence 2208 BP; 601 A; 626 C; 527 G; 454 T; 0 U; 0 Other;
XX
Query Match 83.2%; Score 15.8; DB 14; Length 2208;
Best Local Similarity 89.5%; Pred. No. 1.7e+03;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
QY 1 CTTTACTTCATAGCTTTTG 19
DB 1596 CTTTCTTCATCGCTTTTG 1578

```

```

Pd 14-APR-2005.
XX 30-SEP-2004; 2004WO-US028817.
XX 30-SEP-2003; 2003US-0508226P.
PR 29-APR-2004; 2004US-0566546P.
XX (UYPE-) UNIV PENNSYLVANIA.
XX Wilson JM, Gao G, Alvira MR, Vandenberghe LH;
XX WPI; 2005-285437/29.
XX New adeno-associated virus (AAV) clade comprising at least three AAV
XX members, useful for preventing and/or treating arthritis, multiple
XX sclerosis, diabetes, scleroderma, psoriasis, hemophilia, HIV, bacterial
XX infection and cancer.
XX Disclosure; SEQ ID NO 53; 569pp; English.
XX The invention relates to an adeno-associated virus (AAV) clade comprising
XX at least three AAV members, where each member of the AAV clade is
XX phylogenetically related as determined using a Neighbor-Joining heuristic
XX by a bootstrap value of at least 75 % per 1000 isolates and a Poisson
XX correction distance measurement of no more than 0.05. The methods and
XX compositions of the present invention are useful for the prevention
XX and/or treatment of rheumatoid arthritis, multiple sclerosis,
XX sarcoidosis, diabetes, scleroderma, psoriasis, vasculitis, Crohn's
XX disease, hemophilia, HIV, bacterial infection, cancer and ulcerative
XX colitis. The present sequence represents an adeno-associated virus DNA.
XX
XX Sequence 2208 BP; 602 A; 628 C; 525 G; 453 T; 0 U; 0 Other;
XX
Query Match 83.2%; Score 15.8; DB 14; Length 2208;
Best Local Similarity 89.5%; Pred. No. 1.7e+03;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 1 CTTTACTTCATAGTCCTTTG 19
Db 1596 CTTTCTTCATCGTCCTTTG 1578

RESULT 88
ADZ26901/c
ID ADZ26901 standard; DNA; 2208 BP.
XX AC ADZ26901;
XX 30-JUN-2005 (first entry)
XX Adeno-associated virus DNA SEQ ID NO 51.
XX rheumatoid arthritis; multiple sclerosis; sarcoidosis; diabetes;
XX scleroderma; psoriasis; vasculitis; Crohn's disease; hemophilia; HIV;
XX bacterial infection; cancer; ulcerative colitis; antirheumatic;
XX antiarthritic; neuroprotective; antinflammatory; antidiabetic;
XX antipsoriatic; vasotropic; gastrointestinal-gen.; hemostatic; anti-HIV;
XX virucide; antibacterial; cytostatic; antiulcer; dermatological; ds.
XX Adeno-associated virus.
XX W02005033321-A2.
XX 14-APR-2005.
XX 30-SEP-2004; 2004WO-US028817.
XX 30-SEP-2003; 2003US-0508226P.
XX 29-APR-2004; 2004US-0566546P.
XX (UYPE-) UNIV PENNSYLVANIA.
XX Wilson JM, Gao G, Alvira MR, Vandenberghe LH;
XX WPI; 2005-285437/29.
XX New adeno-associated virus (AAV) clade comprising at least three AAV
XX members, useful for preventing and/or treating arthritis, multiple
XX sclerosis, diabetes, scleroderma, psoriasis, hemophilia, HIV, bacterial
XX infection and cancer.
XX Claim 50; SEQ ID NO 55; 569pp; English.
XX
Query Match 83.2%; Score 15.8; DB 14; Length 2208;
Best Local Similarity 89.5%; Pred. No. 1.7e+03;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 1 CTTTACTTCATAGTCCTTTG 19
Db 1596 CTTTCTTCATCGTCCTTTG 1578

RESULT 89
ADZ26905/c
ID ADZ26905 standard; DNA; 2208 BP.
XX AC ADZ26905;
XX 30-JUN-2005 (first entry)
XX Adeno-associated virus DNA SEQ ID NO 55.
XX rheumatoid arthritis; multiple sclerosis; sarcoidosis; diabetes;
XX scleroderma; psoriasis; vasculitis; Crohn's disease; hemophilia; HIV;
XX bacterial infection; cancer; ulcerative colitis; antirheumatic;
XX antiarthritic; neuroprotective; antinflammatory; antidiabetic;
XX antipsoriatic; vasotropic; gastrointestinal-gen.; hemostatic; anti-HIV;
XX virucide; antibacterial; cytostatic; antiulcer; dermatological; ds.
XX Adeno-associated virus.
XX W02005033321-A2.
XX 14-APR-2005.
XX 30-SEP-2004; 2004WO-US028817.
XX 30-SEP-2003; 2003US-0508226P.
XX 29-APR-2004; 2004US-0566546P.
XX (UYPE-) UNIV PENNSYLVANIA.
XX Wilson JM, Gao G, Alvira MR, Vandenberghe LH;
XX WPI; 2005-285437/29.
XX New adeno-associated virus (AAV) clade comprising at least three AAV
XX members, useful for preventing and/or treating arthritis, multiple
XX sclerosis, diabetes, scleroderma, psoriasis, hemophilia, HIV, bacterial
XX infection and cancer.
XX Claim 50; SEQ ID NO 55; 569pp; English.
XX

```

CC The invention relates to an adeno-associated virus (AAV) clade comprising  
CC at least three AAV members, where each member of the AAV clade is  
CC phylogenetically related as determined using a Neighbor-Joining heuristic  
CC by a bootstrap value of at least 75 % per 1000 isolates and a Poisson  
CC correction distance measurement of no more than 0.05. The methods and  
CC compositions of the present invention are useful for the prevention  
CC and/or treatment of rheumatoid arthritis, multiple sclerosis,  
CC sarcoidosis, diabetes, scleroderma, psoriasis, vasculitis, Crohn's  
CC disease, hemophilia, HIV, bacterial infection, cancer and ulcerative  
CC colitis. The present sequence represents an adeno-associated virus DNA.  
XX  
SQ Sequence 2208 BP; 601 A; 626 C; 526 G; 455 T; 0 U; 0 Other;

Query Match 83.2%; Score 15.8; DB 14; Length 2208;  
Best Local Similarity 89.5%; Pred. No. 1.7e+03;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCCTTTG 19  
|||||  
Db 1596 CTTTCTTCATCGTCCTTTG 1578

RESULT 90  
ADZ26906/c  
ID ADZ26906 standard; DNA; 2208 BP.

XX  
AC ADZ26906;

DT 30-JUN-2005 (first entry)

XX  
DE Adeno-associated virus DNA SEQ ID NO 56.

XX rheumatoid arthritis; multiple sclerosis; sarcoidosis; diabetes;  
XX scleroderma; psoriasis; vasculitis; Crohn's disease; hemophilia; HIV;  
XX bacterial infection; cancer; ulcerative colitis; antirheumatic;  
XX antiarthritic; neuroprotective; antinflammatory; antidiabetic;  
XX antipsoriatic; vasotropic; gastrointestinal-gen.; hemostatic; anti-HIV;  
XX virucide; antibacterial; cytostatic; antitumor; dermatological; ds.

OS Adeno-associated virus.

XX  
PN WO2005033321-A2.

XX  
PD 14-APR-2005.

XX  
PF 30-SEP-2004; 2004WO-US028817.

XX  
PR 30-SEP-2003; 2003US-0508226P.

XX  
PR 29-APR-2004; 2004US-0566546P.

XX  
PA (UYPE-) UNIV PENNSYLVANIA.

XX  
PI Wilson JM, Gao G, Alvira MR, Vandenberghe LH;

XX  
DR WPI; 2005-285437/29.

XX New adeno-associated virus (AAV) clade comprising at least three AAV  
XX members, useful for preventing and/or treating arthritis, multiple  
XX sclerosis, diabetes, scleroderma, psoriasis, hemophilia, HIV, bacterial  
XX infection and cancer.

XX  
PS Claim 50; SEQ ID NO 56; 569pp; English.

XX The invention relates to an adeno-associated virus (AAV) clade comprising  
XX at least three AAV members, where each member of the AAV clade is  
XX phylogenetically related as determined using a Neighbor-Joining heuristic  
XX by a bootstrap value of at least 75 % per 1000 isolates and a Poisson  
XX correction distance measurement of no more than 0.05. The methods and  
XX compositions of the present invention are useful for the prevention  
XX and/or treatment of rheumatoid arthritis, multiple sclerosis,  
XX sarcoidosis, diabetes, scleroderma, psoriasis, vasculitis, Crohn's  
XX disease, hemophilia, HIV, bacterial infection, cancer and ulcerative  
XX colitis. The present sequence represents an adeno-associated virus DNA.

XX  
SQ Sequence 2208 BP; 602 A; 628 C; 527 G; 451 T; 0 U; 0 Other;  
Query Match 83.2%; Score 15.8; DB 14; Length 2208;  
Best Local Similarity 89.5%; Pred. No. 1.7e+03;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCCTTTG 19  
|||||  
Db 1596 CTTTCTTCATCGTCCTTTG 1578

RESULT 91

ADZ26900/c

ID ADZ26900 standard; DNA; 2208 BP.

XX  
AC ADZ26900;

DT 30-JUN-2005 (first entry)

XX  
DE Adeno-associated virus DNA SEQ ID NO 50.

XX rheumatoid arthritis; multiple sclerosis; sarcoidosis; diabetes;  
XX scleroderma; psoriasis; vasculitis; Crohn's disease; hemophilia; HIV;  
XX bacterial infection; cancer; ulcerative colitis; antirheumatic;  
XX antiarthritic; neuroprotective; antinflammatory; antidiabetic;  
XX antipsoriatic; vasotropic; gastrointestinal-gen.; hemostatic; anti-HIV;  
XX virucide; antibacterial; cytostatic; antitumor; dermatological; ds.

OS Adeno-associated virus.

XX  
PN WO2005033321-A2.

XX  
PD 14-APR-2005.

XX  
PF 30-SEP-2004; 2004WO-US028817.

XX  
PR 30-SEP-2003; 2003US-0508226P.

XX  
PR 29-APR-2004; 2004US-0566546P.

XX  
PA (UYPE-) UNIV PENNSYLVANIA.

XX  
PI Wilson JM, Gao G, Alvira MR, Vandenberghe LH;

XX  
DR WPI; 2005-285437/29.

XX New adeno-associated virus (AAV) clade comprising at least three AAV  
XX members, useful for preventing and/or treating arthritis, multiple  
XX sclerosis, diabetes, scleroderma, psoriasis, hemophilia, HIV, bacterial  
XX infection and cancer.

XX  
PS Claim 50; SEQ ID NO 50; 569pp; English.

XX The invention relates to an adeno-associated virus (AAV) clade comprising  
XX at least three AAV members, where each member of the AAV clade is  
XX phylogenetically related as determined using a Neighbor-Joining heuristic  
XX by a bootstrap value of at least 75 % per 1000 isolates and a Poisson  
XX correction distance measurement of no more than 0.05. The methods and  
XX compositions of the present invention are useful for the prevention  
XX and/or treatment of rheumatoid arthritis, multiple sclerosis,  
XX sarcoidosis, diabetes, scleroderma, psoriasis, vasculitis, Crohn's  
XX disease, hemophilia, HIV, bacterial infection, cancer and ulcerative  
XX colitis. The present sequence represents an adeno-associated virus DNA.

XX  
SQ Sequence 2208 BP; 601 A; 627 C; 524 G; 456 T; 0 U; 0 Other;

Query Match 83.2%; Score 15.8; DB 14; Length 2208;  
Best Local Similarity 89.5%; Pred. No. 1.7e+03;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCCTTTG 19  
|||||  
Db 1596 CTTTCTTCATCGTCCTTTG 1578

DE	Adeno-associated virus DNA SEQ ID NO 59.	
XX		
KW	rheumatoid arthritis; multiple sclerosis; sarcoidosis; diabetes;	
KW	scleroderma; psoriasis; vasculitis; Crohn's disease; hemophilia; HIV;	
KW	bacterial infection; cancer; ulcerative colitis; antirheumatic;	
KW	antiarthritic; neuroprotective; antiinflammatory; antidiabetic;	
KW	antipsoriatic; vasotropic; gastrointestinal-gen.; hemostatic; anti-HIV;	
KW	virucide; antibacterial; cytostatic; antitumor; dermatological; ds.	
XX		
OS	Adeno-associated virus.	
XX		
XX	WO2005033321-A2.	
XX		
PD	14-APR-2005.	
XX		
PF	30-SEP-2004; 2004WO-US028817.	
XX		
XX		
PR	20-SEP-2003; 2003US-0508226P.	
PR	29-APR-2004; 2004US-0566546P.	
XX		
XX	(UYPE-) UNIV PENNSYLVANIA.	
PA		
XX		
PI	Wilson JM, Gao G, Alvira MR, Vandenbergh LH;	
XX		
DR	WPI; 2005-285437/29.	
XX		
PT	New adeno-associated virus (AAV) clade comprising at least three AAV	
PT	members, useful for preventing and/or treating arthritis, multiple	
PT	sclerosis, diabetes, scleroderma, psoriasis, hemophilia, HIV, bacterial	
PT	infection and cancer.	
XX		
PS	Disclosure; SEQ ID NO 59; 569pp; English.	
XX		
CC	The invention relates to an adeno-associated virus (AAV) clade comprising	
CC	at least three AAV members, where each member of the AAV clade is	
CC	phylogenetically related as determined using a Neighbor-Joining heuristic	
CC	by a bootstrap value of at least 75 % per 1000 isolates and a Poisson	
CC	correction distance measurement of no more than 0.05. The methods and	
CC	compositions of the present invention are useful for the prevention	
CC	and/or treatment of rheumatoid arthritis, multiple sclerosis,	
CC	sarcoidosis, diabetes, scleroderma, psoriasis, vasculitis, Crohn's	
CC	disease, hemophilia, HIV, bacterial infection, cancer and ulcerative	
CC	colitis. The present sequence represents an adeno-associated virus DNA.	
XX		
SQ	Sequence 2208 BP; 602 A; 627 C; 526 G; 453 T; 0 U; 0 Other;	
Query Match	83.2%; Score 15.8; DB 14; Length 2208;	
Best Local Similarity	89.8%; Pred. No. 1.7e+03;	
Matches	17; Conservative 0; Mismatches 2; Indels/ 0; Gaps 0;	
Qy	1 CTTTACTTCATAGCTTTTG 19	
Dd	1596 CTTTCTTCATCGCTTTTG 1578	
RESULT 94		
ID	ADZ26904/c	
ID	ADZ26904 standard; DNA; 2208 BP.	
XX		
AC	ADZ26904;	
XX		
XX		
DT	30-JUN-2005 (first entry)	
XX		
DE	Adeno-associated virus DNA SEQ ID NO 54.	
XX		
KW	rheumatoid arthritis; multiple sclerosis; sarcoidosis; diabetes;	
KW	scleroderma; psoriasis; vasculitis; Crohn's disease; hemophilia; HIV;	
KW	bacterial infection; cancer; ulcerative colitis; antirheumatic;	
KW	antiarthritic; neuroprotective; antiinflammatory; antidiabetic;	
KW	antipsoriatic; vasotropic; gastrointestinal-gen.; hemostatic; anti-HIV;	
KW	virucide; antibacterial; cytostatic; antitumor; dermatological; ds.	
XX		
OS	Adeno-associated virus.	

RESULT 92  
ADZ26908/c  
ID ADZ26908 standard; DNA; 2208 BP.  
XX AC ADZ26908;  
XX XX  
DT DT 30-JUN-2005 (first entry)  
XX DE Adeno-associated virus DNA SEQ ID NO 58.  
XX KW rheumatoid arthritis; multiple sclerosis; sarcoidosis; diabetes;  
KW KW scleroderma; psoriasis; vasculitis; Crohn's disease; hemophilia; HIV;  
KW KW bacterial infection; cancer; ulcerative colitis; anti-rheumatic;  
KW KW antiarthritic; neuroprotective; antiinflammatory; antidiabetic;  
KW KW antipsoriatic; vasotropic; gastrointestinal-gen.; hemostatic; anti-HIV;  
KW KW virucide; antibacterial; cytostatic; antiulcer; dermatological; ds.  
XX OS Adeno-associated virus.  
XX OS  
XX PN WO2005033321-A2.  
XX XX  
PD PD 14-APR-2005.  
XX PF 30-SEP-2004; 2004WO-US028817.  
XX PR 30-SEP-2003; 2003US-0508226P.  
XX PR 29-APR-2004; 2004US-0566546P.  
XX PA (UYPE-) UNIV PENNSYLVANIA.  
XX PI Wilson JM, Gao G, Alvira MR, Vandenbergh LH;  
XX DR WPI; 2005-285437/29.  
XX XX  
PT PT New adeno-associated virus (AAV) clade comprising at least three AAV  
PT PT members, useful for preventing and/or treating arthritis, multiple  
PT PT sclerosis, diabetes, scleroderma, psoriasis, hemophilia, HIV, bacterial  
PT PT infection and cancer.  
XX XX  
PS PS Claim 7; SEQ ID NO 58; 569pp; English.  
XX CC The invention relates to an adeno-associated virus (AAV) clade comprising  
CC CC at least three AAV members, where each member of the AAV clade is  
CC CC phylogenetically related as determined using a Neighbor-Joining heuristic  
CC CC by a bootstrap value of at least 75 % per 1000 isolates and a Poisson  
CC CC correction distance measurement of no more than 0.05. The methods and  
CC CC compositions of the present invention are useful for the prevention  
CC CC and/or treatment of rheumatoid arthritis, multiple sclerosis,  
CC CC sarcoidosis, diabetes, scleroderma, psoriasis, vasculitis, Crohn's  
CC CC disease, hemophilia, HIV, bacterial infection, cancer and ulcerative  
CC CC colitis. The present sequence represents an adeno-associated virus DNA.

Query Match 83.2%; Score 15.8; DB 14; Length 2208;  
Best Local Similarity 89.5%; Pred. No. 1.7e+03;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTTTACTTCATGCTTTTG 19  
DB 1596 CTTTCTTCATCGCTTTG 1578

RESULT 93  
ADZ26909/c  
ID ADZ26909 standard; DNA; 2208 BP.  
XX AC ADZ26909;  
XX XX  
DT DT 30-JUN-2005 (first entry)  
XX XX

```
XX PN WO2005033321-A2.
XX PD 14-APR-2005.
XX XX
XX PF 30-SEP-2004; 2004WO-US028817.
XX PR 30-SEP-2003; 2003US-0508226P.
XX PR 29-APR-2004; 2004US-0566546P.
XX XX
XX PA (UYPE-) UNIV PENNSYLVANIA.
XX PI Wilson JM, Gao G, Alvira MR, Vandenbergh LH;
XX DR WPI; 2005-285437/29.
XX XX
XX PT New adeno-associated virus (AAV) clade comprising at least three AAV
XX PT members, useful for preventing and/or treating arthritis, multiple
XX PT sclerosis, diabetes, scleroderma, psoriasis, hemophilia, HIV, bacterial
XX PT infection and cancer.
XX PS Disclosure; SEQ ID NO 54; 569pp; English.
XX XX
XX PI Wilson JM, Gao G, Alvira MR, Vandenbergh LH;
XX DR WPI; 2005-285437/29.
XX XX
XX PT New adeno-associated virus (AAV) clade comprising at least three AAV
XX PT members, useful for preventing and/or treating arthritis, multiple
XX PT sclerosis, diabetes, scleroderma, psoriasis, hemophilia, HIV, bacterial
XX PT infection and cancer.
XX PS Disclosure; SEQ ID NO 54; 569pp; English.
XX XX
XX CC The invention relates to an adeno-associated virus (AAV) clade comprising
XX CC at least three AAV members, where each member of the AAV clade is
XX CC phylogenetically related as determined using a Neighbor-Joining heuristic
XX CC by a bootstrap value of at least 75 % per 1000 isolates and a Poisson
XX CC correction distance measurement of no more than 0.05. The methods and
XX CC compositions of the present invention are useful for the prevention
XX CC and/or treatment of rheumatoid arthritis, multiple sclerosis,
XX CC sarcoidosis, diabetes, scleroderma, psoriasis, vasculitis, Crohn's
XX CC disease, hemophilia, HIV, bacterial infection, cancer and ulcerative
XX CC colitis. The present sequence represents an adeno-associated virus DNA.
XX SQ Sequence 2208 BP; 601 A; 626 C; 527 G; 454 T; 0 U; 0 Other;
Query Match 83.2%; Score 15.8; DB 14; Length 2208;
Best Local Similarity 89.5%; Pred. No. 1.7e+03;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTTTACTTCATGCTCTTG 19
Db ||||| ||||| ||||| |||||
1596 CTTTCTTCATGCTCTTG 1578

RESULT 95
ADZ26919/C
ID ADZ26919 standard; DNA; 2208 BP.
XX AC
XX ADZ26919;
XX XX
XX DT 30-JUN-2005 (first entry)
XX DE Adeno-associated virus DNA SEQ ID NO 69.
XX XX
XX KW rheumatoid arthritis; multiple sclerosis; sarcoidosis; diabetes;
XX KW scleroderma; psoriasis; vasculitis; Crohn's disease; hemophilia; HIV;
XX KW bacterial infection; cancer; ulcerative colitis; antirheumatic;
XX KW antiarthritic; neuroprotective; antiinflammatory; antidiabetic;
XX KW antiporiatic; vasotropic; gastrointestinal-gen.; hemostatic; anti-HIV;
XX KW virucide; antibacterial; cytostatic; antiulcer; dermatological; ds.
XX OS
XX PA Adeno-associated virus.
XX XX
XX PN WO2005033321-A2.
XX PD 14-APR-2005.
XX XX
XX PF 30-SEP-2004; 2004WO-US028817.
XX PR 30-SEP-2003; 2003US-0508226P.
XX PR 29-APR-2004; 2004US-0566546P.
XX XX
```

```
PA (UYPE-) UNIV PENNSYLVANIA.
XX PI Wilson JM, Gao G, Alvira MR, Vandenbergh LH;
XX DR WPI; 2005-285437/29.
XX XX
XX PT New adeno-associated virus (AAV) clade comprising at least three AAV
XX PT members, useful for preventing and/or treating arthritis, multiple
XX PT sclerosis, diabetes, scleroderma, psoriasis, hemophilia, HIV, bacterial
XX PT infection and cancer.
XX PS Claim 15; SEQ ID NO 69; 569pp; English.
XX XX
XX CC The invention relates to an adeno-associated virus (AAV) clade comprising
XX CC at least three AAV members, where each member of the AAV clade is
XX CC phylogenetically related as determined using a Neighbor-Joining heuristic
XX CC by a bootstrap value of at least 75 % per 1000 isolates and a Poisson
XX CC correction distance measurement of no more than 0.05. The methods and
XX CC compositions of the present invention are useful for the prevention
XX CC and/or treatment of rheumatoid arthritis, multiple sclerosis,
XX CC sarcoidosis, diabetes, scleroderma, psoriasis, vasculitis, Crohn's
XX CC disease, hemophilia, HIV, bacterial infection, cancer and ulcerative
XX CC colitis. The present sequence represents an adeno-associated virus DNA.
XX SQ Sequence 2208 BP; 589 A; 640 C; 546 G; 433 T; 0 U; 0 Other;
Query Match 83.2%; Score 15.8; DB 14; Length 2208;
Best Local Similarity 89.5%; Pred. No. 1.7e+03;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTTTACTTCATGCTCTTG 19
Db ||||| ||||| ||||| |||||
1596 CTTTCTTCATGCTCTTG 1578

RESULT 96
AAH72833/C
ID AAH72833 standard; cDNA; 2366 BP.
XX AC
XX AAH72833;
XX XX
XX DT 19-SEP-2001 (first entry)
XX DE Human cervical cancer marker nucleic acid 4107.
XX XX
XX KW Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX OS
XX PN WO200142467-A2.
XX PD 14-JUN-2001.
XX XX
XX PF 08-DEC-2000; 2000WO-US033312.
XX PR 08-DEC-1999; 99US-0169681P.
XX PR 21-DEC-1999; 99US-0171350P.
XX PR 14-MAR-2000; 2000US-0189315P.
XX PR 12-MAY-2000; 2000US-0203791P.
XX PR 09-JUN-2000; 2000US-0210600P.
XX PR 21-JUL-2000; 2000US-0220114P.
XX XX
XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX XX
XX PI Schlegel R, Deeds J, Berger A, Zhao X;
XX DR WPI; 2001-375006/39.
XX XX
XX PT New isolated nucleic acid for diagnosing and treating cervical cancer and
XX PT for assessing and detecting compounds for treating the cancer.
XX XX
XX PS Claim 1; Page 877-878; 1051pp; English.
XX XX
```

CC The invention relates to novel genes (AAH68727-AAH73383) associated with  
CC cervical cancer with cytostatic activity. The nucleic acids and encoded  
CC polypeptides are useful: to assess if a patient is afflicted with  
CC cervical cancer or has a pre-malignant condition; to monitor the  
CC progression of cervical cancer or a premalignant condition in a patient;  
CC and to select and/or assess the efficacy of a compound or therapy for  
CC inhibiting cervical cancer in a patient. The nucleic acids may also be  
CC useful for gene therapy

XX SQ Sequence 2366 BP; 746 A; 476 C; 578 G; 559 T; 0 U; 7 Other;

Query Match 83.2%; Score 15.8; DB 4; Length 2366;

Best Local Similarity 89.5%; Pred. No. 1.7e+03;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGCTTTG 19

Db 695 CTTTATTTCATAGCCTTTG 677

RESULT 97

ACC43789/C  
ID ACC43789 standard; DNA; 2376 BP.

XX AC ACC43789;

XX AC 11-AUG-2003 (first entry)

DE Nucleotide sequence of plant regulatory polynucleotide.

XX Plant; regulatory polynucleotide; endothelium; albumen; seed; seed testa;  
KW flavonoid; tannin; cellulose; maize; rape seed; colour; nutrition; ss.

XX Arabidopsis thaliana.

XX Key Location/Qualifiers

FT Promoter 2172..2312

FT CAAT\_signal /\*tag= b

FT TATA\_signal /\*tag= a

FT /\*tag= c

XX FR2828210-A1.

XX 07-FEB-2003.

XX 01-AUG-2001; 2001FR-00010365.

XX 01-AUG-2001; 2001FR-00010365.

XX (INRG ) INRA INST NAT RECH AGRONOMIQUE.

XX (CNRS ) CNRS CENT NAT RECH SCI.

XX Lepiniec L, Debeaujon I, Devic M, Caboche M;

XX WPI; 2003-335321/32.

XX New regulatory nucleic acid, useful for preparing transgenic plants with  
PT altered seed testa, provides specific expression of attached genes in  
PT seed endothelium and albumen.

XX Claim 1; Page 75-76; 85pp; French.

XX The present sequence represents a plant regulatory polynucleotide,  
CC comprising regulatory sequences that allow expression of a polynucleotide  
CC specifically in the endothelium and/or albumen of plant seeds. The  
CC present sequence was used to produce plasmid pBani. The regulatory  
CC polynucleotide, and derived expression cassettes or vectors, are used to  
CC produce transgenic plants in which development of seed testa is altered,  
CC especially for reducing the size or number of seeds, particularly  
CC complete absence of mature, fertile seeds, or modulating contents of  
CC flavonoids, tannins and cellulose (especially to produce maize and rape

CC seeds with less intense yellow colour, or to alter nutritional  
CC properties)

XX SQ Sequence 2376 BP; 804 A; 382 C; 463 G; 727 T; 0 U; 0 Other;

Query Match 83.2%; Score 15.8; DB 8; Length 2376;

Best Local Similarity 89.5%; Pred. No. 1.7e+03;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGCTTTG 19

Db 1040 CTTCACTCCATAGCTTTG 1022

RESULT 98

AAI59130/C

ID AAI59130 standard; cDNA; 2412 BP.

XX AC AAI59130;

XX AC 22-OCT-2001 (first entry)

DE Human polynucleotide SEQ ID NO 1333.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia; ss.

XX Homo sapiens.

XX WO200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US034263.

XX 23-DEC-1999; 99US-00471275.

XX 21-JAN-2000; 2000US-00488725.

XX 25-APR-2000; 2000US-00552317.

XX 20-JUN-2000; 2000US-00598042.

XX 19-JUL-2000; 2000US-00620312.

XX 03-AUG-2000; 2000US-00653450.

XX 14-SEP-2000; 2000US-00662191.

XX 19-OCT-2000; 2000US-00693036.

XX 29-NOV-2000; 2000US-00727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;

XX Zhou P, Goodrich R, Drmanac RT;

XX WPI; 2001-442253/47.

XX P-PSDB; AAM39974.

XX Novel nucleic acids and polypeptides, useful for treating disorders such  
PT as central nervous system injuries.

XX Claim 1; SEQ ID NO 1333; 10078pp; English.

XX The invention relates to human nucleic acids (AAI57798-AAI61369) and the  
CC encoded polypeptides (AAM38642-AAM42213) with nootropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,

CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders. Note: The sequence data for this patent did not form  
CC part of the printed specification

XX SQ Sequence 2412 BP; 793 A; 476 C; 570 G; 573 T; 0 U; 0 Other;

Query Match 83.2%; Score 15.8; DB 4; Length 2412;  
Best Local Similarity 89.5%; Pred. No. 1.7e+03;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCCTTTG 19  
||||| ||||| |||||  
Db 746 CTTTATTTTCATAGCCTTTG 728

RESULT 99  
ADQ99353/c  
ID ADQ99353 standard; cDNA; 2412 BP.  
XX AC ADQ99353;  
XX DT 23-SEP-2004 (first entry)  
XX DE DNA encoding human GPCR-like protein seqid 1023.  
XX KW ophthalmological; immunomodulatory; cytostatic; antiatherosclerotic;  
XX KW antidiabetic; GPCR-like protein; ophthalmic disorder;  
XX KW neurological disorder; immunological disorder; nephritic disorder;  
XX KW hormonal dysfunction; cancer; atherosclerosis; diabetes;  
XX KW molecular weight marker; food supplement; human; ss.  
XX OS Homo sapiens.

XX FN US6569662-B1.

XX PD 27-MAY-2003.

XX PF 19-JUL-2000; 2000US-00620312.

XX PR 21-JAN-2000; 2000US-00488725.

XX PR 25-APR-2000; 2000US-00552317.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Zhou P, Drmanac RT;

XX DR WPI; 2001-442255/47.

XX PT New G-protein-coupled receptor-like polypeptides and polynucleotides,  
XX PT useful for treating diseases of ophthalmic, neurological, immunological  
XX PT and nephritic systems and hormonal dysfunction, cancer, atherosclerosis  
XX PT and diabetes.

XX PS Example 2; SEQ ID NO 1023; 92pp; English.

XX CC The invention describes an isolated polynucleotide (I) comprising a fully  
XX CC defined (S1) of 749, 3188, 2484, 1169, 2936, 1467, 5773, 5714, 4041,  
XX CC 1372, 3996, 3945, 2735, 1788, 585, 1782, 927, 5714 or 2282 nucleotides as  
XX CC given in the specification, its translated or protein coding portion, its  
XX CC extracellular portion or its active domain. The GPCR-like polypeptides  
XX CC and polynucleotides are useful for the treatment of diseases of  
XX CC ophthalmic, neurological, immunological and nephritic systems. They may  
XX CC also be used to treat hormonal dysfunction, cancer, atherosclerosis and  
XX CC diabetes. The antibodies are useful for detecting or quantitating the  
XX CC polypeptide in tissue. The polypeptides can also be used as molecular  
XX CC weight markers and as a food supplement. This sequence represents a human  
XX CC polynucleotide of the invention.

XX SQ Sequence 2412 BP; 793 A; 476 C; 570 G; 573 T; 0 U; 0 Other;

'Query Match 83.2%; Score 15.8; DB 5; Length 2412;

Best Local Similarity 89.5%; Pred. No. 1.7e+03;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCCTTTG 19  
||||| ||||| |||||  
Db 746 CTTTATTTTCATAGCCTTTG 728

RESULT 100  
ADB49113/c  
ID ADB49113 standard; cDNA; 2412 BP.  
XX AC ADB49113;

XX DT 04-DEC-2003 (first entry)

XX DE Novel human cDNA SEQ ID NO 1023.

XX KW ss; cancer; neurodegenerative disease; human.

XX OS Homo sapiens.

XX PN US2003104529-A1.

XX PD 05-JUN-2003.

XX PF 04-JAN-2002; 2002US-00037270.

XX PR 21-JAN-2000; 2000US-00488725.

XX PR 25-APR-2000; 2000US-00552317.

XX PR 19-JUL-2000; 2000US-00620312.

XX PA (ZHOU/) ZHOU P.

XX PA (TANG/) TANG Y T.

XX PA (LIUC/) LIU C.

XX PA (ASUN/) ASUNDI V.

XX PA (DRMA/) DRMANAC R T.

XX PI Zhou P, Tang YT, Liu C, Asundi V, Drmanac RT;

XX DR WPI; 2003-678194/64.

XX PT New polynucleotide, useful for treating diseases e.g., cancer or

XX PT neurodegenerative diseases.

XX PS Claim 1; SEQ ID NO 1023; 99pp; English.

XX CC The invention relates to a polynucleotide comprising a sequence given in  
XX CC the specification, or its mature protein-coding portion, or its  
XX CC complement. The polynucleotide is useful for treating diseases e.g.,  
XX CC cancer or neurodegenerative diseases and many others listed in the  
XX CC specification. The present sequence represents a novel human cDNA. Note:  
XX CC The sequence data for this patent did not form part of the printed  
XX CC specification but was obtained in electronic format directly from USPTO  
XX CC at seqdata.uspto.gov/sequence.html?DocID=20030104529.

XX SQ Sequence 2412 BP; 793 A; 476 C; 570 G; 573 T; 0 U; 0 Other;

Query Match 83.2%; Score 15.8; DB 9; Length 2412;  
Best Local Similarity 89.5%; Pred. No. 1.7e+03;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCCTTTG 19  
||||| ||||| |||||  
Db 746 CTTTATTTTCATAGCCTTTG 728

Search completed: June 10, 2006, 15:32:51  
Job time : 172.483 secs



GenCore version 5.1.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2006, 15:19:28 ; Search time 2344.6 Seconds  
(without alignments)  
453.155 Million cell updates/sec

Title: US-10-600-816-32

Perfect score: 19

Sequence: 1 cttacttcattagctttg 19

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 48236798 seqs, 27959665780 residues

Total number of hits satisfying chosen parameters: 96473596

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

Database :

EST:\*

1: gb\_est1:\*

2: gb\_est3:\*

3: gb\_est4:\*

4: gb\_est5:\*

5: gb\_est6:\*

6: gb\_hic:\*

7: gb\_est2:\*

8: gb\_est7:\*

9: gb\_est8:\*

10: gb\_est9:\*

11: gb\_gss1:\*

12: gb\_gss2:\*

13: gb\_gss3:\*

14: gb\_gss4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	19	100.0	158	2	BI063311 IL3-UT011
C 2	19	100.0	307	7	BE188833 CM2-BN030
C 3	19	100.0	341	7	BE188817 CM2-BN030
C 4	19	100.0	365	7	BE188831 CM2-BN030
C 5	19	100.0	367	1	AA112374 zns6e10.1
C 6	19	100.0	374	7	BE003240 QV4-BN009
C 7	19	100.0	384	7	BE188835 CM2-BN030
C 8	19	100.0	396	7	BE188828 CM2-BN030
C 9	19	100.0	404	7	BE188840 CM2-BN030
C 10	19	100.0	424	7	BE188805 CM2-BN030
C 11	19	100.0	449	7	BE188807 CM2-BN030
C 12	19	100.0	453	5	CD672930 fgl18a05.Y
C 13	19	100.0	458	7	BE188871 CM2-BN030
C 14	19	100.0	466	7	BE188859 CM2-BN030
C 15	19	100.0	467	7	BE188852 CM2-BN030
C 16	19	100.0	571	3	BP253378 BP253378
C 17	19	100.0	577	4	CB152636 K-EST0209
C 18	19	100.0	580	3	BP265234 BP265234
C 19	19	100.0	580	3	BP268151 BP268151

BP256144	BP256144
BP363758	BP363758
BP256147	BP256147
CB130950	K-EST0180
BE747107	601580680
BM843051	K-EST0120
BF568108	602183908
BG251131	602364932
BQ887352	AGENCOURT
CA454746	AGENCOURT
CM011054	603634744
BQ58341	AGENCOURT
BG831564	602765819
BX372160	BX372160
BU543952	AGENCOURT
BQ685729	AGENCOURT
BQ67433	BX370558
CR597125	full-length
DU976902	AUAC-aaa0
DU069153	139562 TO
CF780868	AGENCOURT
BI193620	602946519
DE194703	Branchios
AQ435685	HS 5148 B
AW664308	hi09e10.X
CM050801	v8_p11.p2
BG009082	RC1-GN019
DT490210	WS02544.B
BP267627	BP267627
AM058428	AM058428
AZ353695	1M0092L09
AG158060	Pan trogl
AZ840748	2M0138K24
DV148009	CV03074B2
DT482123	WS02533.B
BU290329	604166187
DN485086	M104E06.3
CV032902	OG0DS21TV
CV242157	WS02514.B
DT506297	WS0189.C2
DT487539	WS02533.B
DT487258	WS02533.B
CV254849	WS024311.B
DT489385	WS02542.B
CV245752	WS0259.B2
DT522272	WS02036.B
DT521741	WS02035.B
CG329079	OG0DS21TH
BE548141	601072396
CG147825	PUIDG50TD
CG330802	OG2BM89TV
CB828264	LJNEST84F
AQ431031	HS 5087 A
CB827270	LJNEST72F
AZ483135	1M0308C10
CT259201	Sus scrofa
BI819265	603034694
DUR46802	34821 Tom
BU870629	99008 Tom
BU931252	AGENCOURT
AK171788	Mus muscu
AK172413	Mus muscu
AK076606	Mus muscu
AK170599	Mus muscu
AK076230	Mus muscu







C 531	15.8	83.2	539	8	CX197198	Sa_mx0_57
C 532	15.8	83.2	540	3	BM601528	170006870
C 533	15.8	83.2	540	3	BM646654	170006873
C 534	15.8	83.2	541	12	CE416906	tigr-g8s-
C 535	15.8	83.2	543	1	AL657096	AL657096
C 536	15.8	83.2	543	1	BM650732	170006873
C 537	15.8	83.2	544	7	AW573670	EST316261
C 538	15.8	83.2	546	3	BM598383	170006876
C 539	15.8	83.2	547	11	AZ025487	RPCI-23-3
C 540	15.8	83.2	549	3	BM621991	170006874
C 541	15.8	83.2	551	3	BM623812	170006874
C 542	15.8	83.2	553	7	AW666072	sk1903.Y
C 543	15.8	83.2	555	7	BF192802	244155.MA
C 544	15.8	83.2	557	12	CN951583	BOID219TF
C 545	15.8	83.2	557	13	DU251279	109853818
C 546	15.8	83.2	558	2	BF971816	sa989g11.
C 547	15.8	83.2	558	4	CB419231	592081.MA
C 548	15.8	83.2	563	5	CD630861	56067738H
C 549	15.8	83.2	563	13	CW760642	OG_BBa006
C 550	15.8	83.2	564	8	CV543492	RTS_135_G
C 551	15.8	83.2	564	11	BH344384	CH230-5P1
C 552	15.8	83.2	566	3	BM614535	170006871
C 553	15.8	83.2	567	3	BP743460	BP743460
C 554	15.8	83.2	568	8	CO493668	G.h.fbr-s
C 555	15.8	83.2	570	6	CNS095YM	AX043802
C 556	15.8	83.2	570	11	BH382339	AG-ND-162
C 557	15.8	83.2	571	8	CO432128	UI-M-HX0-
C 558	15.8	83.2	572	13	CL321226	CH242_15U
C 559	15.8	83.2	576	13	CL712193	OR_BBa003
C 560	15.8	83.2	577	8	CO983527	GM89019A1
C 561	15.8	83.2	577	8	CR441777	CR441777
C 562	15.8	83.2	579	8	CN460669	UI-M-HB0-
C 563	15.8	83.2	580	4	BM269190	BX269190
C 564	15.8	83.2	583	3	BM577967	170006871
C 565	15.8	83.2	583	3	BM646556	170006873
C 566	15.8	83.2	583	8	CN605889	PEL0408_P
C 567	15.8	83.2	583	12	CE248654	tigr-g8s-
C 568	15.8	83.2	585	4	CB438182	686670.MA
C 569	15.8	83.2	585	8	CR417952	CR417952
C 570	15.8	83.2	586	3	BM603160	170006870
C 571	15.8	83.2	589	3	BM598172	170006876
C 572	15.8	83.2	589	3	BU477405	603844352
C 573	15.8	83.2	591	3	BQ856932	QGB6C10.Y
C 574	15.8	83.2	591	6	CNS09FWG	Single re
C 575	15.8	83.2	591	13	CZ675947	ON_Ba023
C 576	15.8	83.2	593	6	CNS09FWH	Single re
C 577	15.8	83.2	593	14	CR911697	Sus scrofa
C 578	15.8	83.2	596	11	AZ222069	RPCI-23-7
C 579	15.8	83.2	598	8	CN669755	A0884A06-
C 580	15.8	83.2	599	12	CE483243	tigr-g8s-
C 581	15.8	83.2	599	14	CR485776	mcH2-154J
C 582	15.8	83.2	600	8	CO429499	UI-M-HB0-
C 583	15.8	83.2	600	11	BH561970	BOGMW41TF
C 584	15.8	83.2	601	11	BH446690	BOGMJ14TR
C 585	15.8	83.2	602	7	AW897906	RC3-NN006
C 586	15.8	83.2	605	1	AL959960	AL959960
C 587	15.8	83.2	605	5	CK655532	AGENCOURT
C 588	15.8	83.2	607	2	BG712561	pglin.pk0
C 589	15.8	83.2	610	13	DU240612	109857417
C 590	15.8	83.2	614	8	CN086155	EC2BBA26D
C 591	15.8	83.2	618	10	DY182842	001016BEM
C 592	15.8	83.2	618	13	CL941999	OA_ABa005
C 593	15.8	83.2	620	13	CL605473	CH240_182
C 594	15.8	83.2	622	5	CF355549	V-B-129D1
C 595	15.8	83.2	622	9	CK736725	797714.MA
C 596	15.8	83.2	623	3	BU703667	UI-M-F00-
C 597	15.8	83.2	623	8	CN076593	EC2BBA12A
C 598	15.8	83.2	623	9	CK736706	796946.MA
C 599	15.8	83.2	623	11	BH382196	AG-ND-105
C 600	15.8	83.2	624	13	CZ636685	OM_Ba018
C 601	15.8	83.2	629	1	AL877874	AL877874
C 602	15.8	83.2	631	10	DV613107	EST121610
C 603	15.8	83.2	632	11	B70649	CIT-HSP-206
CW032361	104	261	1			
AL634018	AL634018					
DN442792	LI85338-0					
CZ660359	OM_Ba021					
CN090189	EC2BBA32D					
CN118959	EC2CAA9BH					
AL677214	AL677214					
AL877353	AL877353					
CA893691	B0180G05-					
CN086312	EC2BBA27D					
CO433246	UI-M-HX0-					
BQ992129	QGF24M13.					
CL334012	RPCI44_25					
CN088569	EC2BBA30B					
BU937287	AGENCOURT					
CX994413	Cm_mx0_06					
CX994340	Cm_mx0_05					
AL857015	AL857015					
AZ995992	2M0282Q06					
BZ871250	CH240_192					
CR966037	mtH4-JN18					
AL652556	AL652556					
CD630862	56067738J					
AL648486	AL648486					
CE791716	tigr-g8s-					
CR561186	AL8671186					
AL879054	CR561186					
BU304357	603733406					
CF534781	UI-M-GH0-					
B1969577	GM830008B					
BJ215782	BJ215782					
CV027639	5946_Full					
BH663938	BOMM28TF					
AL984487	Dario rer					
CD881756	F1_104D02					
AW167183	xg72b08.x					
DR771540	ILLUMIGEN					
CZ112701	OM_Ba015					
CE086063	tigr-g8s-					
BU123649	603147551					
CO243805	WS0036_B2					
CE023679	tigr-g8s-					
BH980752	odf60G10.					
BM957610	fy67c06.Y					
CN535608	UI-M-HS0-					
CK307673	SB02044A2					
BB612388	BB612388					
DM126608	CLSX3150.					
CE083371	tigr-g8s-					
AZ360620	1M0103N23					
CE270717	tigr-g8s-					
AL646912	AL646912					
CR441963	CR441963					
AV823442	AV823442					
BZ337374	ia86h11.b					
CN097884	EC2CAA13C					
BZ223551	CH230-325					
BH966573	odj32F02.					
BZ088431	11a99603.					
CN078014	EC2BBA14B					
CN118958	EC2CAA9BH					
BZ088475	11a99603.					
AL638176	AL638176					
BH276177	CH230-108					
CW694856	OG_BBa005					
CA138605	SCERT202					

677	15.8	83.2	688	14	AG016142	AG016142 Homo sapi	747	1	AL970811	AL970811
678	15.8	83.2	689	10	DT762365	DT762365 EST119621	748	8	CO807632	CO807632 AGENCOURT
679	15.8	83.2	689	11	BH998479	BH998479 ce123h05.	748	9	CX461483	CX461483 JGI_XZG24
680	15.8	83.2	690	11	BH025902	BH025902 RPCT-24-2	748	11	BX081940	BX081940 11e54g06.
681	15.8	83.2	691	14	AG122398	AG122398 Pan trogl	748	14	CR294382	CR294382 mte1-1002
682	15.8	83.2	693	12	CC743712	CC743712 ZMMBB011	749	14	CN455376	CN455376 UI-M-HN0-
683	15.8	83.2	695	8	CV177655	CV177655 LEFISNA45	749	12	CG859361	CG859361 ZMMBSC026
684	15.8	83.2	696	3	BU208321	BU208321 603104166	750	4	BW729019	BW729019 BW729019
685	15.8	83.2	698	9	CX496277	CX496277 JGI_XZG38	750	4	CA765776	CA765776 AF53-Rpf
686	15.8	83.2	698	9	DN267104	DN267104 L1B30320	750	8	CR587686	CR587686 CR587686
687	15.8	83.2	699	13	CW566004	CW566004 OA_Aba009	751	9	DN025001	DN025001 JGI_CAAAR4
688	15.8	83.2	700	3	BM633530	BM633530 17006875	751	9	DN025002	DN025002 JGI_CAAAR4
689	15.8	83.2	700	11	AZ258992	AZ258992 RPCT-23-1	752	8	CR417953	CR417953 CR417953
690	15.8	83.2	701	12	CC518998	CC518998 CH240_365	753	4	BF775165	BF775165 BX775165
691	15.8	83.2	702	11	BW759219	BW759219 BW759219	753	5	CF783611	CF783611 AGENCOURT
692	15.8	83.2	702	11	BH936769	BH936769 odg52f08.	753	11	AG117637	AG117637 Pan trogl
693	15.8	83.2	702	14	AG178050	AG178050 Pan trogl	754	14	AG117637	AG117637 Pan trogl
694	15.8	83.2	703	12	BZ654887	BZ654887 OGANE607M	755	8	CN086154	CN086154 EC2BBA26D
695	15.8	83.2	705	8	CN459027	CN459027 UI-M-HN0-	756	10	DV578484	DV578484 0061F0024
696	15.8	83.2	705	10	DY370977	DY370977 ZO_Ef000	756	10	DV578485	DV578485 0061F0024
697	15.8	83.2	705	11	AZ553100	AZ553100 RPCT-23-2	757	8	CN078013	CN078013 EC2BBA14B
698	15.8	83.2	707	9	DN049795	DN049795 JGI_CABA5	757	8	CN088568	CN088568 EC2BBA30B
699	15.8	83.2	707	14	AG321751	AG321751 Mus muscu	757	8	CV271341	CV271341 WS0154.B2
700	15.8	83.2	708	8	CN108252	CN108252 EC2CAR31C	757	9	CX343624	CX343624 JGI_XZT22
701	15.8	83.2	708	9	DN049796	DN049796 JGI_CABA5	757	9	CX880346	CX880346 JGI_CAAAL2
702	15.8	83.2	708	12	CL182780	CL182780 104_394_1	757	14	AG525039	AG525039 Mus muscu
703	15.8	83.2	709	5	CF783250	CF783250 AGENCOURT	758	8	CN536767	CN536767 UI-M-HS0-
704	15.8	83.2	710	10	DV188880	DV188880 BCT032_G0	758	10	DT428498	DT428498 JGI_CABJ7
705	15.8	83.2	711	11	BZ028663	BZ028663 oef01a11.	759	4	BX753374	BX753374 BX753374
706	15.8	83.2	712	14	CT142332	CT142332 Sus scrofa	759	9	DN084603	DN084603 JGI_CABD1
707	15.8	83.2	714	5	CD351692	CD351692 UI-M-G10-	759	9	DN084603	DN084603 JGI_CABD1
708	15.8	83.2	715	14	CR346329	CR346329 mte1-8N15	759	10	DT428499	DT428499 JGI_CABJ7
709	15.8	83.2	715	13	DU454641	DU454641 109841579	759	10	DT453795	DT453795 JGI_CABK1
710	15.8	83.2	715	5	CF749061	CF749061 UI-M-HJ0-	760	11	BH141740	BH141740 TDGDH17TH
711	15.8	83.2	715	8	CN086911	CN086911 EC2BBA27D	762	5	CJ477741	CJ477741 CJ477741
712	15.8	83.2	715	14	CT268937	CT268937 Sus scrofa	762	11	BZ081909	BZ081909 11e54g06.
713	15.8	83.2	716	8	CN535770	CN535770 UI-M-HS0-	765	8	CV271616	CV271616 WS0155.B2
714	15.8	83.2	717	5	CK654491	CK654491 AGENCOURT	767	8	CR562925	CR562925 CR562925
715	15.8	83.2	717	8	CR586444	CR586444 CR586444	768	4	BX731259	BX731259 BX731259
716	15.8	83.2	717	12	BZ654879	BZ654879 OGANE607C	769	1	AJ787440	AJ787440 AJ787440
717	15.8	83.2	719	5	CK307034	CK307034 SB02042A2	769	14	DR835499	DR835499 JGI_CABC5
718	15.8	83.2	719	13	CL814711	CL814711 OR_CBA003	770	9	DN020250	DN020250 JGI_CAAAR2
719	15.8	83.2	720	14	AG515277	AG515277 Mus muscu	770	12	CC786231	CC786231 ZMMBB0015
720	15.8	83.2	721	8	CN901018	CN901018 EC2BBA32D	773	8	CN308840	CN308840 328769055
721	15.8	83.2	721	14	DU575742	DU575742 OO_Ba006	775	11	BZ059780	BZ059780 11931f04.
722	15.8	83.2	721	14	DU599467	DU599467 OO_Ba009	776	4	BX745148	BX745148 BX745148
723	15.8	83.2	723	10	DU530632	DU530632 UI-M-HO0-	778	9	DN084604	DN084604 JGI_CABD1
724	15.8	83.2	723	10	DY353628	DY353628 ZO_Ec000	778	11	BZ120200	BZ120200 CH230-411
725	15.8	83.2	724	4	BX731260	BX731260 BX731260	779	12	CC584051	CC584051 CH240_380
726	15.8	83.2	724	5	CK640202	CK640202 UI-M-HN0-	782	9	CX343625	CX343625 JGI_XZT22
727	15.8	83.2	727	8	CN533965	CN533965 UI-M-HO0-	782	12	CC756467	CC756467 ZMMBB0014
728	15.8	83.2	727	9	CX998994	CX998994 JGI_CAAQ2	783	8	CR580586	CR580586 CR580586
729	15.8	83.2	729	9	AZ551772	AZ551772 RPCT-23-2	783	9	CX496278	CX496278 JGI_XZG38
730	15.8	83.2	730	13	CL836108	CL836108 OR_CBA006	783	13	DU209662	DU209662 109857429
731	15.8	83.2	732	11	BH982803	BH982803 odF13g09.	784	8	CR425988	CR425988 CR425988
732	15.8	83.2	732	13	CZ359508	CZ359508 ZMMBF0100	784	10	DT453796	DT453796 JGI_CABK1
733	15.8	83.2	735	4	BX696593	BX696593 BX696593	785	5	CF151627	CF151627 AGENCOURT
734	15.8	83.2	735	14	DX162770	DX162770 OR_Aba001	785	5	CJ452997	CJ452997 CJ452997
735	15.8	83.2	736	8	CR441778	CR441778 CR441778	785	11	BH263821	BH263821 CH230-161
736	15.8	83.2	737	8	CR596445	CR596445 CR596445	786	8	CX232101	CX232101 MBM05936
737	15.8	83.2	737	8	CN097883	CN097883 EC2CAAL3C	786	14	CT351904	CT351904 Sus scrofa
738	15.8	83.2	739	14	DU599419	DU599419 OO_Ba009	788	12	CC629216	CC629216 OGCA0909TH
739	15.8	83.2	740	8	CR580587	CR580587 CR580587	790	3	BU121255	BU121255 603144866
740	15.8	83.2	740	8	CX226795	CX226795 MBM02315	790	5	CF709562	CF709562 CCAF324TR
741	15.8	83.2	740	13	CZ614878	CZ614878 OM_Ba014	790	8	CR562924	CR562924 CR562924
742	15.8	83.2	742	8	CX312500	CX312500 JGI_XZT85	790	10	DV578476	DV578476 0061P0015
743	15.8	83.2	742	14	CR181844	CR181844 Reverse 6	790	10	DV578477	DV578477 0061P0015
744	15.8	83.2	743	9	CX998993	CX998993 JGI_CAAQ2	791	4	BX703018	BX703018 BX703018
745	15.8	83.2	743	9	DN017894	DN017894 JGI_CAAAR7	792	4	BX706436	BX706436 BX706436
746	15.8	83.2	744	11	BZ031497	BZ031497 oefh34h03.	792	5	CJ474837	CJ474837 CJ474837
747	15.8	83.2	746	8	NO76592	NO76592 EC2BBA12A	793	4	BX703019	BX703019 BX703019
748	15.8	83.2	746	10	DV574714	DV574714 0058P0015	793	10	DV578482	DV578482 0061P0012
749	15.8	83.2	746	10	DV574715	DV574715 0058P0015	793	10	DV578483	DV578483 0061P0012



```
c 969 15..8 83..2 950 7 BE311673
970 15..8 83..2 950 14 DX043737
971 15..8 83..2 952 11 BH157372
972 15..8 83..2 953 14 C9S06HSY
973 15..8 83..2 965 14 AG033435
974 15..8 83..2 968 4 BY716162
975 15..8 83..2 973 13 CL444650
976 15..8 83..2 974 3 BU514398
977 15..8 83..2 974 5 CF242598
978 15..8 83..2 984 1 AV337438
979 15..8 83..2 990 6 AY223165
980 15..8 83..2 1002 8 CV861525
981 15..8 83..2 1004 1 AL568980
982 15..8 83..2 1005 1 AM094865
983 15..8 83..2 1021 1 AM094863
984 15..8 83..2 1028 12 BZ699902
985 15..8 83..2 1033 13 CW928899
986 15..8 83..2 1036 13 CL235772
987 15..8 83..2 1057 3 BU482478
988 15..8 83..2 1060 12 CC243593
989 15..8 83..2 1064 14 AY404891
990 15..8 83..2 1066 13 CL455059
991 15..8 83..2 1068 14 C9S05856
992 15..8 83..2 1081 13 C2399014
993 15..8 83..2 1100 12 CC248668
994 15..8 83..2 1101 14 AL104876
995 15..8 83..2 1104 13 C2965649
996 15..8 83..2 1115 14 AG388201
997 15..8 83..2 1116 13 CL948053
998 15..8 83..2 1119 12 CL021010
999 15..8 83..2 1120 12 CC224627
1000 15..8 83..2 1180 11 BZ550135

ALIGNMENTS

BI063311 158 bp mRNA linear EST 15-JUN-2001
IL3-UT0117-300301-538-E01 UT0117 Homo sapiens cDNA, mRNA sequence.
BI063311
VERSION EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.
1 (bases 1 to 158)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
10737800
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL3&t2=IL3-UT0117-
300301-538-E01&t3=2001-03-30&t4=1)
Seq primer: puc 18 forward

FEATURES
source
Location/Qualifiers
1..158
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="UT0117"
/note="Organ: uterus_tumor; Vector: puc18; Site:1: SmaI;
Site 2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
ORIGIN
Query Match 100.0%; Score 19; DB 2; Length 158;
Best Local Similarity 100.0%; Pred. No. 3e+02; 0; Indels 0; Gaps 0;
Matches 19; Conservative 0; Mismatches 0;
1 CTTTACTTCATAGTCTTTG 19
|||||
141 CTTTACTTCATAGTCTTTG 123

RESULT 2
BE188833 307 bp mRNA linear EST 21-SEP-2000
CM2-BN0302-050700-256-c01 BN0302 Homo sapiens cDNA, mRNA sequence.
BE188833
VERSION EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.
1 (bases 1 to 307)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
10737800
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM2-BN0302-050
700-256-c01&t3=2000-07-05&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 25
High quality sequence stop: 307.
High quality sequence stop: 307.
Location/Qualifiers
1..307
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="BN0302"
/note="Organ: breast normal; Vector: puc18; Site 1: SmaI;
Site 2: SmaI; A mini-library was made by cloning products
```



derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

## ORIGIN

Query Match 100.0%; Score 19; DB 7; Length 307;  
Best Local Similarity 100.0%; Pred. No. 3.2e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19  
|||||  
Db 201 CTTTACTTCATAGTCTTTG 219

## RESULT 3

BE818817  
LOCUS BE818817 341 bp mRNA linear EST 21-SEP-2000  
DEFINITION CM2-BN0302-050700-256-b07 BN0302 Homo sapiens cDNA, mRNA sequence.

ACCESSION BE818817  
VERSION BE818817.1 GI:10251051  
KEYWORDS EST.  
SOURCE Homo sapiens (human)

## ORGANISM

Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.

## REFERENCE

1 (bases 1 to 341)  
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Mateukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F.F., Brentani, R.R., Reis, L.F.F., de Souza, S.J. and Simpson, A.J.J.  
Shotgun sequencing of the human transcriptome with ORF expressed

## TITLE

Shotgun sequencing of the human transcriptome with ORF expressed

## JOURNAL

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

## PUBMED

## COMMENT

Contact: Simpson A.J.J.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2=CM2-BN0302-050  
700-256-b07&t3=2000-07-05&t4=1)

Seq primer: puc 18 forward  
High quality sequence start: 10  
High quality sequence stop: 341.

## FEATURES

## source

1..341  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/dev\_stage="Adult"  
/clone\_lib="BN0302"  
/note="Organ: breast normal; Vector: puc18; Site 1: SmaI;  
Site 2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

## ORIGIN

Query Match 100.0%; Score 19; DB 7; Length 341;  
Best Local Similarity 100.0%; Pred. No. 3.2e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19  
|||||  
Db 212 CTTTACTTCATAGTCTTTG 230

## RESULT 4

BE818831  
LOCUS BE818831 365 bp mRNA linear EST 21-SEP-2000  
DEFINITION CM2-BN0302-050700-256-a12 BN0302 Homo sapiens cDNA, mRNA sequence.

ACCESSION BE818831  
VERSION BE818831.1 GI:10251065  
KEYWORDS EST.  
SOURCE Homo sapiens (human)

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.

## REFERENCE

1 (bases 1 to 365)  
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Mateukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F.F., Brentani, R.R., Reis, L.F.F., de Souza, S.J. and Simpson, A.J.J.  
Shotgun sequencing of the human transcriptome with ORF expressed

## sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

## PUBMED

## COMMENT

Contact: Simpson A.J.J.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2=CM2-BN0302-050  
700-256-a12&t3=2000-07-05&t4=1)

Seq primer: puc 18 forward  
High quality sequence start: 14  
High quality sequence stop: 365.

## FEATURES

## source

1..365  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/dev\_stage="Adult"  
/clone\_lib="BN0302"  
/note="Organ: breast normal; Vector: puc18; Site 1: SmaI;  
Site 2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

## ORIGIN

Query Match 100.0%; Score 19; DB 7; Length 365;  
Best Local Similarity 100.0%; Pred. No. 3.3e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19  
|||||  
Db 226 CTTTACTTCATAGTCTTTG 244

## RESULT 5

AA112374/c  
LOCUS AA112374 367 bp mRNA linear EST 01-AUG-1997  
DEFINITION zn68e10.r1 Stratagene HeLa cell s3 937216 Homo sapiens cDNA clone



Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: [asimpson@ludwig.org.br](mailto:asimpson@ludwig.org.br)  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2-CM2-BN0302-050>)  
700-256-cltkt3=2000-07-05&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 5  
High quality sequence stop: 384.

## FEATURES

source  
1..384  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/dev\_stage="Adult"  
/clone\_lib="BN0302"  
/notes="Organ: breast normal; Vector: puc18; Site: 1: SmaI;  
Site 2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

## ORIGIN

Query Match 100.0%; Score 19; DB 7; Length 384;  
Best Local Similarity 100.0%; Pred. No. 3.3e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19  
|||||

Db 216 CTTTACTTCATAGTCTTTG 234  
|||||

## RESULT 8

BE818828  
LOCUS BE818828 396 bp mRNA linear EST 21-SEP-2000  
DEFINITION CM2-BN0302-050700-256-a05 BN0302 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BE818828  
VERSION BE818828.1 GI:10251062  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
1 (bases 1 to 396)

## REFERENCE

AUTHORS  
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.,  
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,  
Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V.,  
O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
Simpson, A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

## PUBMED

10737800

## COMMENT

Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: [asimpson@ludwig.org.br](mailto:asimpson@ludwig.org.br)

This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2-CM2-BN0302-050>)

700-256-a05&t3=2000-07-05&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 11  
High quality sequence stop: 396.

## FEATURES

source  
1..396  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/dev\_stage="Adult"  
/clone\_lib="BN0302"  
/notes="Organ: breast normal; Vector: puc18; Site: 1: SmaI;  
Site 2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

## ORIGIN

Query Match 100.0%; Score 19; DB 7; Length 396;  
Best Local Similarity 100.0%; Pred. No. 3.3e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19  
|||||

Db 207 CTTTACTTCATAGTCTTTG 225  
|||||

## RESULT 9

BE818840  
LOCUS BE818840 404 bp mRNA linear EST 21-SEP-2000  
DEFINITION CM2-BN0302-050700-256-f02 BN0302 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BE818840  
VERSION BE818840.1 GI:10251074  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
1 (bases 1 to 404)

## REFERENCE

AUTHORS  
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.,  
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,  
Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V.,  
O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
Simpson, A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

## PUBMED

10737800

## COMMENT

Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: [asimpson@ludwig.org.br](mailto:asimpson@ludwig.org.br)

This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2-CM2-BN0302-050>)  
700-256-f02&t3=2000-07-05&t4=1)

Seq primer: puc 18 forward

High quality sequence start: 27

High quality sequence stop: 404.

## FEATURES

source  
1..404  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/dev\_stage="Adult"

/clone\_lib="BN0302"  
 /notes="Organ: breast normal; Vector: puc18; Site\_1: SmaI;  
 Site\_2: SmaI; A mini-library was made by cloning products  
 derived from ORESTES PCR (U.S. Letters Patent application  
 No. 196,716 - Ludwig Institute for Cancer Research)  
 profiles into the pUC 18 vector. Reverse transcription of  
 tissue mRNA and cDNA amplification were performed under  
 low stringency conditions."

## ORIGIN

Query Match 100.0%; Score 19; DB 7; Length 404;  
 Best Local Similarity 100.0%; Pred. No. 3.3e+02;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCTTTG 19  
 |||||  
 Db 203 CTTTACTTCATAGTCTTTG 221

## RESULT 10

BE818805  
 LOCUS 424 bp mRNA linear EST 21-SEP-2000  
 DEFINITION CM2-BN0302-040700-253-b10 BN0302 Homo sapiens cDNA, mRNA sequence.  
 ACCESSION BE818805  
 VERSION BE818805.1 GI:10251039  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homnidae; Homo.  
 1 (bases 1 to 424)

REFERENCE  
AUTHORS

Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
 Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,  
 Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,  
 O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
 Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

## JOURNAL

## PUBMED

## COMMENT

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome  
 Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2=CM2-BN0302-040  
 700-253-b10&t3=2000-07-04&t4=1)

Seq primer: puc 18 forward

High quality sequence start: 4

High quality sequence stop: 424.

## FEATURES

## source

1..424  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /dev\_stage="Adult"  
 /clone\_lib="BN0302"

/notes="Organ: breast normal; Vector: puc18; Site\_1: SmaI;  
 Site\_2: SmaI; A mini-library was made by cloning products  
 derived from ORESTES PCR (U.S. Letters Patent application  
 No. 196,716 - Ludwig Institute for Cancer Research)  
 profiles into the pUC 18 vector. Reverse transcription of  
 tissue mRNA and cDNA amplification were performed under  
 low stringency conditions."

## ORIGIN

Query Match 100.0%; Score 19; DB 7; Length 424;  
 Best Local Similarity 100.0%; Pred. No. 3.3e+02;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCTTTG 19  
 |||||  
 Db 209 CTTTACTTCATAGTCTTTG 227

## RESULT 11

## BE818807

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## 1 (bases 1 to 449)

## Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,

## Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,

## Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,

## Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,

## O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and

## Simpson,A.J.

## Shotgun sequencing of the human transcriptome with ORF expressed

## sequence tags

## Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

## JOURNAL

## PUBMED

## COMMENT

## Contact: Simpson A.J.G.

## Laboratory of Cancer Genetics

## Ludwig Institute for Cancer Research

## Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

## Brazil

## Tel: +55-11-2704922

## Fax: +55-11-2707001

## Email: asimpson@ludwig.org.br

## This sequence was derived from the FAPESP/LICR Human Cancer Genome

## Project. This entry can be seen in the following URL

## (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2=CM2-BN0302-040

## 700-253-d02&amp;t3=2000-07-04&amp;t4=1)

## Seq primer: puc 18 forward

## High quality sequence start: 24

## High quality sequence stop: 306.

## FEATURES

## Location/Qualifiers

## 1..449

## /organism="Homo sapiens"

## /mol\_type="mRNA"

## /db\_xref="taxon:9606"

## /dev\_stage="Adult"

## /clone\_lib="BN0302"

## /notes="Organ: breast normal; Vector: puc18; Site\_1: SmaI;

## Site\_2: SmaI; A mini-library was made by cloning products

## derived from ORESTES PCR (U.S. Letters Patent application

## No. 196,716 - Ludwig Institute for Cancer Research)

## profiles into the pUC 18 vector. Reverse transcription of

## tissue mRNA and cDNA amplification were performed under

## low stringency conditions."

## ORIGIN

## Query Match

## Best Local Similarity

## Matches

## QY

## Db

## RESULT 12

CD672930/c  
 LOCUS fgl18a05.y2 Human Iris cDNA (Normalized): fg Homo sapiens cDNA clone  
 fgl18a05.5', mRNA sequence.  
 ACCESSION CD672930  
 VERSION CD672930.1 GI:32174661  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 453)  
 AUTHORS Wistow,G., Bernstein,S.L., Ray,S., Wyatt,M.K., Behal,A., Touchman,J.W., Bouffard,G., Smith,D. and Peterson,K.  
 TITLE Expressed sequence tag analysis of adult human iris for the NEIBank Project: steroid-response factors and similarities with retinal pigment epithelium  
 JOURNAL Mol. Vis. 8 (4), 185-195 (2002)  
 PUBMED 12107412  
 COMMENT Contact: Wistow G  
 Section on Molecular Structure and Function  
 National Eye Institute  
 6/331, NIH, Bethesda, MD 20892-2740, USA  
 Tel: 301 402 3452  
 Fax: 301 496 0078  
 Email: graeme@helix.nih.gov  
 Plate: 18 row: a column: 05  
 Seq primer: M13RP1 reverse primer (ABI).  
 Location/Qualifiers  
 1. 453  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="fgl18a05"  
 /tissue\_type="Iris"  
 /dev\_stage="Adult"  
 /lab\_host="EMDH10B"  
 /clone\_lib="Human Iris cDNA (Normalized): fg"  
 /notes="Organ: Eye; Vector: pCMVSPORT6; A human iris library (bx) was normalized by self-subtraction. One portion of double stranded plasmid DNA representing the library was linearized by NotI. This NotI digested library was used as a template for biotinylated RNA synthesis using SP6 RNA polymerase. Another portion of the double stranded plasmid library was converted to single-stranded circles in vitro using Gene II and Exonuclease III (Life Technologies). Single-stranded DNA (1 mg) was hybridized (Cot 500) with 41 mg of Bio-RNA and vector blocking oligonucleotides. The hybridized Bio-RNA/ss-circles were removed by streptavidin:phenol extraction. EST analysis was performed on the library at the NIH Intramural Sequencing Center (NISC)."

FEATURES  
 source  
 1. 453  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="fgl18a05"  
 /tissue\_type="Iris"  
 /dev\_stage="Adult"  
 /lab\_host="EMDH10B"  
 /clone\_lib="Human Iris cDNA (Normalized): fg"  
 /notes="Organ: Eye; Vector: pCMVSPORT6; A human iris library (bx) was normalized by self-subtraction. One portion of double stranded plasmid DNA representing the library was linearized by NotI. This NotI digested library was used as a template for biotinylated RNA synthesis using SP6 RNA polymerase. Another portion of the double stranded plasmid library was converted to single-stranded circles in vitro using Gene II and Exonuclease III (Life Technologies). Single-stranded DNA (1 mg) was hybridized (Cot 500) with 41 mg of Bio-RNA and vector blocking oligonucleotides. The hybridized Bio-RNA/ss-circles were removed by streptavidin:phenol extraction. EST analysis was performed on the library at the NIH Intramural Sequencing Center (NISC)."

ORIGIN  
 Query Match 100.0%; Score 19; DB 5; Length 453;  
 Best Local Similarity 100.0%; Pred. No. 3.3e+02;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CTTTACTTCATAGTCTTTG 19  
 |||||  
 Db 97 CTTTACTTCATAGTCTTTG 79

RESULT 13  
 BE818871  
 LOCUS CM2-BN0302-100700-259-f01 BN0302 Homo sapiens cDNA, mRNA sequence.  
 DEFINITION BE818871  
 ACCESSION BE818871  
 VERSION BE818871.1 GI:10251105  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 458)  
 AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.  
 TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 PUBMED 10737800  
 COMMENT Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2=CM2-BN0302-100700-259-f01&tl3=2000-07-10&tl4=1)  
 Seq primer: puc 18 forward  
 High quality sequence start: 12  
 High quality sequence stop: 458.

FEATURES  
 source  
 1. 458  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /dev\_stage="Adult"  
 /clone\_lib="BN0302"  
 /notes="Organ: breast normal; Vector: puc18; Site: 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from OSESTES PCR (U.S. Letters patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN  
 Query Match 100.0%; Score 19; DB 7; Length 458;  
 Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CTTTACTTCATAGTCTTTG 19  
 |||||  
 Db 223 CTTTACTTCATAGTCTTTG 241

RESULT 14  
 BE818859  
 LOCUS CM2-BN0302-100700-259-c02 BN0302 Homo sapiens cDNA, mRNA sequence.  
 DEFINITION BE818859  
 ACCESSION BE818859.1 GI:10251093  
 VERSION EST.  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 466)  
 AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 458)  
 Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.  
 TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 PUBMED 10737800  
 COMMENT Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2=CM2-BN0302-100700-259-f01&tl3=2000-07-10&tl4=1)  
 Seq primer: puc 18 forward  
 High quality sequence start: 12  
 High quality sequence stop: 458.

FEATURES  
 source  
 1. 458  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /dev\_stage="Adult"  
 /clone\_lib="BN0302"  
 /notes="Organ: breast normal; Vector: puc18; Site: 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from OSESTES PCR (U.S. Letters patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN  
 Query Match 100.0%; Score 19; DB 7; Length 458;  
 Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CTTTACTTCATAGTCTTTG 19  
 |||||  
 Db 223 CTTTACTTCATAGTCTTTG 241

RESULT 14  
 BE818859  
 LOCUS CM2-BN0302-100700-259-c02 BN0302 Homo sapiens cDNA, mRNA sequence.  
 DEFINITION BE818859  
 ACCESSION BE818859.1 GI:10251093  
 VERSION EST.  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 466)  
 AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

**TITLE** Shotgun sequencing of the human transcriptome with ORF expressed  
**JOURNAL** sequence tags  
**PUBLISHED** Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
**COMMENT** 10737800

**CONTACT:** Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=&t2=CM2-BN0302-100  
700-259-c02&t3=2000-07-10&t4=1)

Seq primer: puc 18 forward  
High quality sequence start: 16  
High quality sequence stop: 466.

**FEATURES**  
source  
1..466  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/dev\_stage="Adult"  
/clone\_lib="BN0302"  
/notes="Organ: breast normal; Vector: puc18; Site 1: SmaI;  
Site 2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

**ORIGIN**  
Query Match 100.0%; Score 19; DB 7; Length 466;  
Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

**QY** 1 CTTTACTTCATAGTCCTTTG 19  
|||||  
**Db** 226 CTTTACTTCATAGTCCTTTG 244

**RESULT 15**  
**BE818852**  
**LOCUS** BE818852 467 bp mRNA linear EST 21-SEP-2000  
**DEFINITION** CM2-BN0302-100700-259-a05 BN0302 Homo sapiens cDNA, mRNA sequence.  
**ACCESSION** BE818852  
**VERSION** BE818852.1 GI:10251086  
**KEYWORDS** EST.  
**SOURCE** Homo sapiens (human)  
**ORGANISM** Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
1 (bases 1 to 467)  
Dias Neto,E., Garcia Correa,R., Vexjovski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,  
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,  
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.

**TITLE** Shotgun sequencing of the human transcriptome with ORF expressed  
**JOURNAL** sequence tags  
**PUBLISHED** Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
**COMMENT** 10737800

**CONTACT:** Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922

**Fax:** +55-11-2707001  
**Email:** asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=&t2=CM2-BN0302-100  
700-259-a05&t3=2000-07-10&t4=1)

Seq primer: puc 18 forward  
High quality sequence start: 19  
High quality sequence stop: 467.

**FEATURES**  
Location/Qualifiers  
1..467  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/dev\_stage="Adult"  
/clone\_lib="BN0302"  
/notes="Organ: breast normal; Vector: puc18; Site 1: SmaI;  
Site 2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

**ORIGIN**  
Query Match 100.0%; Score 19; DB 7; Length 467;  
Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

**QY** 1 CTTTACTTCATAGTCCTTTG 19  
|||||  
**Db** 205 CTTTACTTCATAGTCCTTTG 223

**RESULT 16**  
**BP253378/c**  
**LOCUS** BP253378 571 bp mRNA linear EST 15-SEP-2004  
**DEFINITION** BP253378 Sugano cDNA library, Kidney epithelial cell Homo sapiens  
cDNA clone HRC03720, mRNA sequence.  
**ACCESSION** BP253378  
**VERSION** BP253378.1 GI:52135659  
**KEYWORDS** EST.  
**SOURCE** Homo sapiens (human)  
**ORGANISM** Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
1 (bases 1 to 571)  
Suzuki,X., Yamashita,R., Shiota,M., Sakakibara,Y., Chiba,J.,  
Mizushima-Sugano,J., Nakai,K. and Sugano,S.  
Sequence comparison of human and mouse genes reveals a homologous  
block structure in the promoter regions  
Genome Res. 14 (9), 1711-1718 (2004)  
15342556  
Contact: Yutaka Suzuki  
Department of Virology  
Institute of Medical Science, University of Tokyo  
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
Tel: 81-3-5449-5343  
Fax: 81-3-5449-5416  
Email: yusuzuki@ncc.jp.

**FEATURES**  
Location/Qualifiers  
1..571  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone\_lib="HRC03720"  
/tissue\_type="kidney"  
/cell\_type="epithelial cell"  
/clone\_lib="Sugano cDNA library, kidney epithelial cell"

**ORIGIN**  
Query Match 100.0%; Score 19; DB 3; Length 571;

Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCTTTG 19  
|||||  
Db 563 CTTTACTTCATAGTCTTTG 545

RESULT 17  
CB152636/c  
LOCUS CB152636 577 bp mRNA linear EST 29-JAN-2003  
DEFINITION K-EST0209868 L12JSHC0s1 Homo sapiens cDNA clone L12JSHC0s1-5-G07  
5', mRNA sequence.

ACCESSION CB152636  
VERSION CB152636.1 GI:28137590  
KEYWORDS EST.

SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.

1 (bases 1 to 577)  
Kim, N.S., Hahn, Y., Oh, J.H., Lee, J.Y., Ahn, H.Y., Chu, M.Y., Kim, M.R.,  
Oh, K.J., Cheong, J.E., Sohn, H.Y., Kim, J.M., Park, H.S., Kim, S. and  
Kim, Y.S.

21C Frontier Korean EST Project 2001  
Unpublished (2002)

CONTACT: Kim YS

Genome Research Center  
Korea Research Institute of Bioscience & Biotechnology  
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea

Tel: +82-42-860-4470  
Fax: +82-42-860-4409

Email: yongsung@mail.kribb.re.kr

Plate: 5, row: G column: 07

High quality sequence stop: 577.

Location/Qualifiers

1. 577

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="L12JSHC0s1-5-G07"

/sex="M"

/cell\_line="J-SHC"

/lab\_host="Top10P"

/clone\_lib="L12JSHC0s1"

/notes="Organ: Liver; Vector: pCNS-D2; Site 1: EcoRI;  
Site 2: NotI; The poly (A) + RNA was dephosphorylated with

bacterial alkaline phosphatase (BAP) and then decapped  
with tobacco acid pyrophosphatase (TAP). The decapped

intact mRNA was ligated with DNA-RNA linker including  
EcoRI site by treatment of T4 RNA ligase and the first

strand cDNA was synthesized from oligo dt-selected mRNA by  
priming with dt-tailed vector. The dt-tailed vector was

adjusted to have about 60nt. The cDNA vector was  
circularized with E. coli DNA ligase after digestion of

EcoRI which site is also included in vector. An RNA strand  
converted to a DNA strand by Okayama-Berg method. The

obtained cDNA vectors were used for transformation of  
competent cells E. coli Top10P' by electroporation method.

The cDNA libraries constructed by this method are  
full-length enriched cDNA library. After analyzing and

sequencing about 2,000 - 3,000 colonies in original cDNA  
library, the abundant cDNAs were selected and amplified by

PCR reaction using vector region primer including T7  
promotor as 5' primer and N(dt)14 as 3' primer. The PCR

products were used as template for synthesis of  
biotinylated single stranded RNA by in vitro transcription

reaction. The synthesized RNA probes were hybridized with  
antisense single stranded cDNAs prepared from original

library and incubated with avidin-gel. After removing  
DNA-RNA hybrids by centrifuge, the subtracted cDNA

libraries were constructed by transfection of the

# ORIGIN

Query Match 100.0%; Score 19; DB 4; Length 577;  
Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCTTTG 19  
|||||  
Db 238 CTTTACTTCATAGTCTTTG 220

RESULT 18  
BP265234/c

LOCUS BP265234 580 bp mRNA linear EST 16-SEP-2004  
DEFINITION BP265234 Sugano cDNA library, thyroid JTH Homo sapiens cDNA clone  
JTH00322, mRNA sequence.

ACCESSION BP265234

VERSION BP265234.1 GI:52180465

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.

1 (bases 1 to 580)

Suzuki, Y., Yamashita, R., Shiota, M., Sakakibara, Y., Chiba, J.,

Mizushima-Sugano, J., Nakai, K. and Sugano, S.

Sequence comparison of human and mouse genes reveals a homologous

block structure in the promoter regions

Genome Res. 14 (9), 1711-1718 (2004)

15342556

Contact: Yutaka Suzuki

Department of Virology

Institute of Medical Science, University of Tokyo

4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan

Tel: 81-3-5449-5343

Fax: 81-3-5449-5416

Email: yuzuki@hgc.jp

Location/Qualifiers

1. 580

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="JTH00322"

/tissue\_type="thyroid"

/cell\_line="JTH"

/clone\_lib="Sugano cDNA library, thyroid JTH"

/note="thyroid tumor"

# ORIGIN

Query Match 100.0%; Score 19; DB 3; Length 580;  
Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCTTTG 19  
|||||  
Db 368 CTTTACTTCATAGTCTTTG 350

RESULT 19  
BP268151/c

LOCUS BP268151 580 bp mRNA linear EST 16-SEP-2004  
DEFINITION BP268151 Sugano cDNA library, thyroid JTH Homo sapiens cDNA clone  
JTH09144, mRNA sequence.

ACCESSION BP268151

VERSION BP268151.1 GI:52183383

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

```
REFERENCE
AUTHORS      Hominidae; Homo.
              1 (bases 1 to 580)
              Suzuki,Y., Yamashita,R., Shiota,M., Sakakibara,Y., Chiba,J.,
              Mizushima-Sugano,J., Nakai,K. and Sugano,S.
TITLE        Sequence comparison of human and mouse genes reveals a homologous
JOURNAL      block structure in the promoter regions
PUBMED       Genome Res. 14 (9), 1711-1718 (2004)
COMMENT      15342556
              Contact: Yutaka Suzuki
              Department of Virology
              Institute of Medical Science, University of Tokyo
              4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
              Tel: 81-3-5449-5343
              Fax: 81-3-5449-5416
              Email: ysuzuki@hgc.jp.
FEATURES
source
1..580
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="JTH09144"
/tissue_type="thyroid"
/cell_line="JTH"
/clone_lib="Sugano cDNA library, thyroid JTH"
/note="thyroid tumor"
ORIGIN
Query Match      100.0%; Score 19; DB 3; Length 580;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 CTTTACTTCATAGTCCTTTG 19
Db      579 CTTTACTTCATAGTCCTTTG 561

RESULT 20
BP256144/c
LOCUS      BP256144 Sugano cDNA library, kidney epithelial cell Homo sapiens
DEFINITION      cDNA clone HRC10671, mRNA sequence.
ACCESSION      BP256144
VERSION        BP256144.1 GI:52171374
KEYWORDS       EST.
SOURCE         Homo sapiens (human)
ORGANISM       Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
               Hominidae; Homo.
REFERENCE      1 (bases 1 to 582)
AUTHORS        Suzuki,Y., Yamashita,R., Shiota,M., Sakakibara,Y., Chiba,J.,
               Mizushima-Sugano,J., Nakai,K. and Sugano,S.
TITLE          Sequence comparison of human and mouse genes reveals a homologous
JOURNAL        block structure in the promoter regions
PUBMED         Genome Res. 14 (9), 1711-1718 (2004)
COMMENT        15342556
               Contact: Yutaka Suzuki
               Department of Virology
               Institute of Medical Science, University of Tokyo
               4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
               Tel: 81-3-5449-5343
               Fax: 81-3-5449-5416
               Email: ysuzuki@hgc.jp.
FEATURES
source
1..582
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="HRC10671"
/tissue_type="kidney"
/cell_type="epithelial cell"
/clone_lib="Sugano cDNA library, kidney epithelial cell"
ORIGIN
Query Match      100.0%; Score 19; DB 3; Length 580;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 CTTTACTTCATAGTCCTTTG 19
Db      579 CTTTACTTCATAGTCCTTTG 561

RESULT 20
BP256144/c
LOCUS      BP256144 Sugano cDNA library, kidney epithelial cell Homo sapiens
DEFINITION      cDNA clone HRC10671, mRNA sequence.
ACCESSION      BP256144
VERSION        BP256144.1 GI:52171374
KEYWORDS       EST.
SOURCE         Homo sapiens (human)
ORGANISM       Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
               Hominidae; Homo.
REFERENCE      1 (bases 1 to 582)
AUTHORS        Suzuki,Y., Yamashita,R., Shiota,M., Sakakibara,Y., Chiba,J.,
               Mizushima-Sugano,J., Nakai,K. and Sugano,S.
TITLE          Sequence comparison of human and mouse genes reveals a homologous
JOURNAL        block structure in the promoter regions
PUBMED         Genome Res. 14 (9), 1711-1718 (2004)
COMMENT        15342556
               Contact: Yutaka Suzuki
               Department of Virology
               Institute of Medical Science, University of Tokyo
               4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
               Tel: 81-3-5449-5343
               Fax: 81-3-5449-5416
               Email: ysuzuki@hgc.jp.
FEATURES
source
1..582
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="HRC10671"
/tissue_type="kidney"
/cell_type="epithelial cell"
/clone_lib="Sugano cDNA library, kidney epithelial cell"
ORIGIN
Query Match      100.0%; Score 19; DB 3; Length 582;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 CTTTACTTCATAGTCCTTTG 19
Db      421 CTTTACTTCATAGTCCTTTG 403

RESULT 22
BP256147/c
LOCUS      BP256147 Sugano cDNA library, fetal lung fibroblast TIG Homo
DEFINITION      cDNA clone HRC10676, mRNA sequence.
ACCESSION      BP256147
VERSION        BP256147.1 GI:52171377
KEYWORDS       EST.
SOURCE         Homo sapiens (human)
ORGANISM       Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
               Hominidae; Homo.
REFERENCE      1 (bases 1 to 582)
AUTHORS        Suzuki,Y., Yamashita,R., Shiota,M., Sakakibara,Y., Chiba,J.,
               Mizushima-Sugano,J., Nakai,K. and Sugano,S.
TITLE          Sequence comparison of human and mouse genes reveals a homologous
JOURNAL        block structure in the promoter regions
PUBMED         Genome Res. 14 (9), 1711-1718 (2004)
COMMENT        15342556
               Contact: Yutaka Suzuki
               Department of Virology
               Institute of Medical Science, University of Tokyo
               4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
               Tel: 81-3-5449-5343
               Fax: 81-3-5449-5416
               Email: ysuzuki@hgc.jp.
FEATURES
source
1..582
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="HRC10676"
/tissue_type="lung"
/cell_type="fibroblast"
/cell_line="TIG"
/dev_stage="fetal"
/clone_lib="Sugano cDNA library, fetal lung fibroblast
TIG"
ORIGIN
Query Match      100.0%; Score 19; DB 3; Length 582;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 CTTTACTTCATAGTCCTTTG 19
Db      421 CTTTACTTCATAGTCCTTTG 403

RESULT 22
BP256147/c
LOCUS      BP256147 Sugano cDNA library, kidney epithelial cell Homo sapiens
DEFINITION      cDNA clone HRC10676, mRNA sequence.
ACCESSION      BP256147
VERSION        BP256147.1 GI:52171377
KEYWORDS       EST.
SOURCE         Homo sapiens (human)
ORGANISM       Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
               Hominidae; Homo.
REFERENCE      1 (bases 1 to 582)
AUTHORS        Suzuki,Y., Yamashita,R., Shiota,M., Sakakibara,Y., Chiba,J.,
               Mizushima-Sugano,J., Nakai,K. and Sugano,S.
TITLE          Sequence comparison of human and mouse genes reveals a homologous
JOURNAL        block structure in the promoter regions
PUBMED         Genome Res. 14 (9), 1711-1718 (2004)
COMMENT        15342556
               Contact: Yutaka Suzuki
               Department of Virology
               Institute of Medical Science, University of Tokyo
               4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
               Tel: 81-3-5449-5343
               Fax: 81-3-5449-5416
               Email: ysuzuki@hgc.jp.
FEATURES
source
1..582
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="HRC10676"
/tissue_type="lung"
/cell_type="fibroblast"
/cell_line="TIG"
/dev_stage="fetal"
/clone_lib="Sugano cDNA library, fetal lung fibroblast
TIG"
ORIGIN
Query Match      100.0%; Score 19; DB 3; Length 582;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 CTTTACTTCATAGTCCTTTG 19
Db      421 CTTTACTTCATAGTCCTTTG 403

RESULT 22
BP256147/c
LOCUS      BP256147 Sugano cDNA library, kidney epithelial cell Homo sapiens
DEFINITION      cDNA clone HRC10676, mRNA sequence.
ACCESSION      BP256147
VERSION        BP256147.1 GI:52171377
KEYWORDS       EST.
SOURCE         Homo sapiens (human)
ORGANISM       Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
               Hominidae; Homo.
REFERENCE      1 (bases 1 to 582)
AUTHORS        Suzuki,Y., Yamashita,R., Shiota,M., Sakakibara,Y., Chiba,J.,
               Mizushima-Sugano,J., Nakai,K. and Sugano,S.
TITLE          Sequence comparison of human and mouse genes reveals a homologous
JOURNAL        block structure in the promoter regions
PUBMED         Genome Res. 14 (9), 1711-1718 (2004)
COMMENT        15342556
               Contact: Yutaka Suzuki
               Department of Virology
               Institute of Medical Science, University of Tokyo
               4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
               Tel: 81-3-5449-5343
               Fax: 81-3-5449-5416
               Email: ysuzuki@hgc.jp.
FEATURES
source
1..582
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="HRC10676"
/tissue_type="lung"
/cell_type="fibroblast"
/cell_line="TIG"
/dev_stage="fetal"
/clone_lib="Sugano cDNA library, fetal lung fibroblast
TIG"
ORIGIN
Query Match      100.0%; Score 19; DB 3; Length 582;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 CTTTACTTCATAGTCCTTTG 19
Db      421 CTTTACTTCATAGTCCTTTG 403

RESULT 22
BP256147/c
LOCUS      BP256147 Sugano cDNA library, fetal lung fibroblast TIG Homo
DEFINITION      sapiens cDNA clone TIR02383, mRNA sequence.
ACCESSION      BP363758
VERSION        BP363758.1 GI:52293963
KEYWORDS       EST.
SOURCE         Homo sapiens (human)
ORGANISM       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
               Hominidae; Homo.
REFERENCE      1 (bases 1 to 582)
AUTHORS        Suzuki,Y., Yamashita,R., Shiota,M., Sakakibara,Y., Chiba,J.,
               Mizushima-Sugano,J., Nakai,K. and Sugano,S.
TITLE          Sequence comparison of human and mouse genes reveals a homologous
JOURNAL        block structure in the promoter regions
PUBMED         Genome Res. 14 (9), 1711-1718 (2004)
COMMENT        15342556
               Contact: Yutaka Suzuki
               Department of Virology
               Institute of Medical Science, University of Tokyo
               4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
               Tel: 81-3-5449-5343
               Fax: 81-3-5449-5416
               Email: ysuzuki@hgc.jp.
FEATURES
source
1..582
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="TIR02383"
/tissue_type="lung"
/cell_type="fibroblast"
/cell_line="TIG"
/dev_stage="fetal"
/clone_lib="Sugano cDNA library, fetal lung fibroblast
TIG"
ORIGIN
Query Match      100.0%; Score 19; DB 3; Length 582;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 CTTTACTTCATAGTCCTTTG 19
Db      456 CTTTACTTCATAGTCCTTTG 438

RESULT 21
BP363758/c
LOCUS      BP363758 Sugano cDNA library, fetal lung fibroblast TIG Homo
DEFINITION      sapiens cDNA clone TIR02383, mRNA sequence.
ACCESSION      BP363758
VERSION        BP363758.1 GI:52293963
KEYWORDS       EST.
SOURCE         Homo sapiens (human)
ORGANISM       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
               Hominidae; Homo.
REFERENCE      1 (bases 1 to 582)
AUTHORS        Suzuki,Y., Yamashita,R., Shiota,M., Sakakibara,Y., Chiba,J.,
               Mizushima-Sugano,J., Nakai,K. and Sugano,S.
TITLE          Sequence comparison of human and mouse genes reveals a homologous
JOURNAL        block structure in the promoter regions
PUBMED         Genome Res. 14 (9), 1711-1718 (2004)
COMMENT        15342556
               Contact: Yutaka Suzuki
               Department of Virology
               Institute of Medical Science, University of Tokyo
               4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
               Tel: 81-3-5449-5343
               Fax: 81-3-5449-5416
               Email: ysuzuki@hgc.jp.
FEATURES
source
1..582
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="TIR02383"
/tissue_type="lung"
/cell_type="fibroblast"
/cell_line="TIG"
/dev_stage="fetal"
/clone_lib="Sugano cDNA library, fetal lung fibroblast
TIG"
ORIGIN
Query Match      100.0%; Score 19; DB 3; Length 582;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 CTTTACTTCATAGTCCTTTG 19
Db      456 CTTTACTTCATAGTCCTTTG 438
```



```

REFERENCE
AUTHORS      1 (bases 1 to 584)
              Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
              Mizushano-Sugano,J., Nakai,K. and Sugano,S.
TITLE        Sequence comparison of human and mouse genes reveals a homologous
              block structure in the promoter regions
JOURNAL      Genome Res. 14 (9), 1711-1718 (2004)
PUBMED       15342556
COMMENT      Contact: Yutaka Suzuki
              Department of Virology
              Institute of Medical Science, University of Tokyo
              4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
              Tel: 81-3-5449-5343
              Fax: 81-3-5449-5416
              Email: yuzuki@bqc.jp.
FEATURES     Location/Qualifiers
source       1..584
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"
              /clone="HRC10676"
              /tissue_type="kidney"
              /cell_type="epithelial cell"
              /clone_lib="Sugano cDNA library, kidney epithelial cell"

ORIGIN
Query Match      100.0%; Score 19; DB 3; Length 584;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGCTTTG 19
    |||||
Db 457 CTTTACTTCATAGCTTTG 439

RESULT 23
CB130950/c
LOCUS       CB130950
DEFINITION K-EST0180946 L12JSHCO Homo sapiens cDNA clone L12JSHCO-5-D03 5',
            mRNA sequence.
ACCESSION  CB130950
VERSION     CB130950.1 GI:28095412
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1 (bases 1 to 621)
            Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
            Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
            Kim,Y.S.
            21C Frontier Korean EST Project 2001
            Unpublished (2002)
            Contact: Kim YS
            Genome Research Center
            Korea Research Institute of Bioscience & Biotechnology
            52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
            Tel: +82-42-860-4470
            Fax: +82-42-860-4409
            Email: yongsung@mail.kribb.re.kr
            Plate: 5 row: D column: 03
            High quality sequence stop: 621.
            Location/Qualifiers
FEATURES     source
              1..621
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"
              /clone="L12JSHCO-5-D03"
              /sex="M"
              /cell_line="J-SHC"
              /lab_host="Top10F"
              /clone_lib="L12JSHCO"
              /note="Organ: Liver; Vector: pcNS-D2; Site_1: EcoRI;

```

```

              Site_2: NotI; The poly (A) + RNA was dephosphorylated with
              bacterial alkaline phosphatase (BAP) and then decapped
              with tobacco acid pyrophosphatase (TAP). The decapped
              intact mRNA was ligated with DNA-RNA linker including
              EcoRI site by treatment of T4 RNA ligase and the first
              strand cDNA was synthesized from oligo dt-selected mRNA by
              priming with dt-tailed vector. The dt-tailed vector was
              adjusted to have about 60nt. The cDNA vector was
              circularized with E. coli DNA ligase after digestion of
              EcoRI which site is also included in vector. An RNA strand
              converted to a DNA strand by Okayama-Berg method. The
              obtained cDNA vectors were used for transformation of
              competent cells E. coli Top10F by electroporation method.
              The cDNA libraries constructed by this method are
              full-length enriched cDNA library."

ORIGIN
Query Match      100.0%; Score 19; DB 4; Length 621;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGCTTTG 19
    |||||
Db 355 CTTTACTTCATAGCTTTG 337

RESULT 24
BE747107/c
LOCUS       BE747107
DEFINITION 601580680F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3929706 5',
            mRNA sequence.
ACCESSION  BE747107
VERSION     BE747107.1 GI:10161099
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1 (bases 1 to 630)
            NIH-MGC http://mgc.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgabs-x@mail.nih.gov
            Tissue procurement: DCTD/DTF
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LCM762 row: o column: 19
            High quality sequence stop: 628.
            Location/Qualifiers
FEATURES     source
              1..630
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"
              /clone="IMAGE:3929706"
              /tissue_type="adenocarcinoma cell line"
              /lab_host="DH10B (phage-resistant)"
              /clone_lib="NIH_MGC_9"
              /note="Organ: ovary; Vector: pOTB7; Site 1: XhoI; Site 2:
              EcoRI; cDNA made by oligo-dT priming. Directionally
              cloned into EcoRI/XhoI sites using the following 5',
              adaptor: GGCACGAG(G). Size-selected >500bp for average
              insert size 1.8kb. Library constructed by Ling Hong in
              the laboratory of Gerald M. Rubin (University of
              California, Berkeley) using ZAP-cDNA synthesis kit
              (Stratagene) and Superscript II RT (Life Technologies)."

ORIGIN
Query Match      100.0%; Score 19; DB 7; Length 630;

```

Best Local Similarity 100.0%; Pred. No. 3.5e+02; Mismatches 0; Indels 0; Gaps 0;  
Matches 19; Conservative 0;

QY 1 CTTTACTTCATAGTCTTTG 19  
|||||  
Db 107 CTTTACTTCATAGTCTTTG 89  
|||||

RESULT 25  
BM843051/c  
LOCUS  
DEFINITION K-EST0120649 S12SNU216 Homo sapiens cDNA clone S12SNU216-77-H12 5',  
mRNA sequence.

ACCESSION BM843051  
VERSION BM843051.1 GI:19199460  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 637)  
AUTHORS Oh, J.H., Yang, J.O., Hahn, Y., Kim, M.R., Byun, S.S., Jeon, Y.J., Kim, J.M., Song, K.S., Noh, S.M., Kim, S., Yoo, H.S., Kim, Y.S. and Kim, N.S.  
TITLE Transcriptome analysis of human gastric cancer  
JOURNAL Mamm. Genome 16 (12), 942-954 (2005)  
PUBMED 16341674  
COMMENT Contact: Kim YS  
Genome Research Center  
Korea Research Institute of Bioscience & Biotechnology  
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea  
Tel: +82-42-860-4470  
Fax: +82-42-860-4409  
Email: yongseung@mail.kribb.re.kr  
Plate: 77 row: H column: 12  
High quality sequence stop: 637.  
Location/Qualifiers  
1..637

FEATURES  
source  
1..637  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="S12SNU216-77-H12"  
/sex="F"  
/tissue\_type="Lymph node"  
/cell\_line="Epithelial"  
/lab\_host="Top10F"  
/clone\_lib="S12SNU216"  
/notes="Organ: Stomach; Vector: pCNS; Site 1: EcoRI; Site 2: NotI; The poly (A)+ RNA was dephosphorylated with bacterial alkaline phosphatase (BAP) and then decapped with tabacco acid pyrophosphatase (TAP). The decapped intact mRNA was ligated with DNA-RNA linker including EcoR I site by treatment of T4 RNA ligase and the first strand cDNA was synthesized from oligo dt-selected mRNA by priming with dt-tailed vector. The dt-tailed vector was adjusted to have about 60nt. The cDNA vector was circularized with E. coli DNA ligase after digestion of EcoRI which site is also included in vector. An RNA strand converted to a DNA strand by Okayama-Berg method. The obtained cDNA vectors were used for transformation of competent cells E. coli Top10F by electroporation method. The cDNA libraries constructed by this method are full-length enriched cDNA library."

ORIGIN  
Query Match 100.0%; Score 19; DB 3; Length 637;  
Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCTTTG 19  
|||||

Db 100 CTTTACTTCATAGTCTTTG 82

RESULT 26  
BF568108/c  
LOCUS  
DEFINITION BF568108F1 NIH\_MGC\_42 Homo sapiens cDNA clone IMAGE:4300248 5',  
mRNA sequence.

ACCESSION BF568108  
VERSION BF568108.1 GI:11641526  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 699)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabs-r@mail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L1CM158 row: O column: 01  
High quality sequence stop: 638.  
Location/Qualifiers  
1..699  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4300248"  
/tissue\_type="epithelioid carcinoma cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC 42"  
/notes="Organ: pancreas; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

FEATURES  
source  
1..699  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4300248"  
/tissue\_type="epithelioid carcinoma cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC 42"  
/notes="Organ: pancreas; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

ORIGIN  
Query Match 100.0%; Score 19; DB 7; Length 699;  
Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCTTTG 19  
|||||  
Db 362 CTTTACTTCATAGTCTTTG 344  
|||||

RESULT 27  
BG251131/c  
LOCUS  
DEFINITION BG251131F1 NIH\_MGC\_90 Homo sapiens cDNA clone IMAGE:4473439 5',  
mRNA sequence.

ACCESSION BG251131  
VERSION BG251131.1 GI:12760947  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 702)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabs-r@mail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L1CM158 row: O column: 01  
High quality sequence stop: 638.  
Location/Qualifiers  
1..702  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4300248"  
/tissue\_type="epithelioid carcinoma cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC 42"  
/notes="Organ: pancreas; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

FEATURES  
source  
1..702  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4300248"  
/tissue\_type="epithelioid carcinoma cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC 42"  
/notes="Organ: pancreas; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

ORIGIN  
Query Match 100.0%; Score 19; DB 7; Length 699;  
Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCTTTG 19  
|||||  
Db 362 CTTTACTTCATAGTCTTTG 344  
|||||

RESULT 27  
BG251131/c  
LOCUS  
DEFINITION BG251131F1 NIH\_MGC\_90 Homo sapiens cDNA clone IMAGE:4473439 5',  
mRNA sequence.

ACCESSION BG251131  
VERSION BG251131.1 GI:12760947  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 702)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabs-r@mail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L1CM158 row: O column: 01  
High quality sequence stop: 638.  
Location/Qualifiers  
1..702  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4300248"  
/tissue\_type="epithelioid carcinoma cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC 42"  
/notes="Organ: pancreas; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

FEATURES  
source  
1..702  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4300248"  
/tissue\_type="epithelioid carcinoma cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC 42"  
/notes="Organ: pancreas; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

ORIGIN  
Query Match 100.0%; Score 19; DB 7; Length 699;  
Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCTTTG 19  
|||||  
Db 362 CTTTACTTCATAGTCTTTG 344  
|||||

RESULT 27  
BG251131/c  
LOCUS  
DEFINITION BG251131F1 NIH\_MGC\_90 Homo sapiens cDNA clone IMAGE:4473439 5',  
mRNA sequence.

ACCESSION BG251131  
VERSION BG251131.1 GI:12760947  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 702)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabs-r@mail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L1CM158 row: O column: 01  
High quality sequence stop: 638.  
Location/Qualifiers  
1..702  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4300248"  
/tissue\_type="epithelioid carcinoma cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC 42"  
/notes="Organ: pancreas; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

FEATURES  
source  
1..702  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4300248"  
/tissue\_type="epithelioid carcinoma cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC 42"  
/notes="Organ: pancreas; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

ORIGIN  
Query Match 100.0%; Score 19; DB 7; Length 699;  
Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCTTTG 19  
|||||  
Db 362 CTTTACTTCATAGTCTTTG 344  
|||||

RESULT 27  
BG251131/c  
LOCUS  
DEFINITION BG251131F1 NIH\_MGC\_90 Homo sapiens cDNA clone IMAGE:4473439 5',  
mRNA sequence.

ACCESSION BG251131  
VERSION BG251131.1 GI:12760947  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 702)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabs-r@mail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L1CM158 row: O column: 01  
High quality sequence stop: 638.  
Location/Qualifiers  
1..702  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4300248"  
/tissue\_type="epithelioid carcinoma cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC 42"  
/notes="Organ: pancreas; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

FEATURES  
source  
1..702  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4300248"  
/tissue\_type="epithelioid carcinoma cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC 42"  
/notes="Organ: pancreas; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

ORIGIN  
Query Match 100.0%; Score 19; DB 7; Length 699;  
Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCTTTG 19  
|||||

RESULT 27  
BG251131/c  
LOCUS  
DEFINITION BG251131F1 NIH\_MGC\_90 Homo sapiens cDNA clone IMAGE:4473439 5',  
mRNA sequence.

ACCESSION BG251131  
VERSION BG251131.1 GI:12760947  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 702)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabs-r@mail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L1CM158 row: O column: 01  
High quality sequence stop: 638.  
Location/Qualifiers  
1..702  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4300248"  
/tissue\_type="epithelioid carcinoma cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC 42"  
/notes="Organ: pancreas; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

```

REFERENCE
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM10295 row: 0 column: 08
High quality sequence stop: 688.
FEATURES
source Location/Qualifiers
1. .702
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4473439"
/tissue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_90"
/note="Organ: liver; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.7 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
ORIGIN
Query Match 100.0%; Score 19; DB 2; Length 702;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 CTTTACTTCATAGTCTTTG 19
|||||
Db 505 CTTTACTTCATAGTCTTTG 487

RESULT 28
BQ887352/c
LOCUS BQ887352 851 bp mRNA linear EST 16-AUG-2002
DEFINITION AGENCOURT 8678064 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6381080
5', mRNA sequence.
ACCESSION BQ887352
VERSION BQ887352.1 GI:22279366
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1 (bases 1 to 851)
NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/DTF
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM2570 row: 1 column: 09
High quality sequence stop: 579.
FEATURES
source Location/Qualifiers
1. .851
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"

```

```

/clone="IMAGE:6381080"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_40"
/note="Organ: prostate; Vector: pOTB7; Site 1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
ORIGIN
Query Match 100.0%; Score 19; DB 3; Length 851;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 CTTTACTTCATAGTCTTTG 19
|||||
Db 223 CTTTACTTCATAGTCTTTG 205

RESULT 29
CA454746/c
LOCUS CA454746 865 bp mRNA linear EST 12-NOV-2002
DEFINITION AGENCOURT_10763191 MAPcL Homo sapiens cDNA clone IMAGE:6721245 5',
mRNA sequence.
ACCESSION CA454746
VERSION CA454746.1 GI:24904781
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1 (bases 1 to 865)
NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Kristi A. Eglund, Ira Pastan
cDNA Library Preparation: Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM14282 row: a column: 21
High quality sequence stop: 686.
FEATURES
source Location/Qualifiers
1. .865
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6721245"
/cell_line="ZR-75-1, MCF7, SK-BR-3, MDA-MB-231,
hTERT-HME1, LNCaP"
/lab_host="EMDH10B"
/clone_lib="MAPcL"
/note="Vector: pCMV-SPORT6; Site 1: EcoRV; Site 2: Not I;
Subtracted with brain, liver, lung, kidney and muscle.
Directionally cloned. Priming method: oligo-dT. Average
insert size: 1800 bp. Library amplification: 26,000 fold.
Kristi A. Eglund, James J. Vincent, Robert Strausberg,
Bungkook Lee & Ira Pastan. Discovery of new breast
cancer genes encoding membrane and secreted proteins.
Manuscript submitted."
ORIGIN
Query Match 100.0%; Score 19; DB 4; Length 865;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;

```

```

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCCTTTG 19
|||||
Db 652 CTTTACTTCATAGTCCTTTG 634

RESULT 30
BM011054/c
LOCUS
DEFINITION
  BM011054 877 bp mRNA linear EST 30-OCT-2001
  603634744F1 NIH_MGC_47 Homo sapiens cDNA clone IMAGE:5421755 5',
  mRNA sequence.
ACCESSION
  BM011054
VERSION
  BM011054.1 GI:16525408
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
  Homnidae; Homo.
REFERENCE
  1 (bases 1 to 877)
  NIH-MGC http://mgs.nci.nih.gov/.
  National Institutes of Health, Mammalian Gene Collection (MGC)
  Unpublished (1999)
  Contact: Robert Strausberg, Ph.D.
  Email: cgabbs-r@mail.nih.gov
  Tissue Procurement: ATCC
  cDNA Library Preparation: Ling Hong/Rubin Laboratory
  DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
  DNA Sequencing by: Incyte Genomics, Inc.
  Clone distribution: MGC clone distribution information can be
  found through the I.M.A.G.E. Consortium/LLNL at:
  http://image.llnl.gov
  Plate: L1CM1878 row: h column: 12
  High quality sequence stop: 746.
  Location/Qualifiers
    1..877
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone="IMAGE:5421755"
    /tissue_type="neuroblastoma, cell line"
    /lab_host="DH10B (phage-resistant)"
    /clone_lib="NIH MGC 47"
    /notes="Organ: brain; Vector: pOTB7; Site 1: XhoI; Site 2:
    EcoRI; cDNA made by oligo-dt priming. Directionally
    cloned into EcoRI/XhoI sites using the following 5',
    adaptor: GGCACGAG(G). Size-selected >500bp for average
    insert size 1.8kb. Library constructed by Ling Hong in
    the laboratory of Gerald M. Rubin (University of
    California, Berkeley) using ZAP-cDNA synthesis kit
    (Stratagene) and Superscript II RT (Life Technologies).
    Note: this is a NIH_MGC Library."

FEATURES
  source
    1..877
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone="IMAGE:5421755"
    /tissue_type="neuroblastoma, cell line"
    /lab_host="DH10B (phage-resistant)"
    /clone_lib="NIH MGC 47"
    /notes="Organ: brain; Vector: pOTB7; Site 1: XhoI; Site 2:
    EcoRI; cDNA made by oligo-dt priming. Directionally
    cloned into EcoRI/XhoI sites using the following 5',
    adaptor: GGCACGAG(G). Size-selected >500bp for average
    insert size 1.8kb. Library constructed by Ling Hong in
    the laboratory of Gerald M. Rubin (University of
    California, Berkeley) using ZAP-cDNA synthesis kit
    (Stratagene) and Superscript II RT (Life Technologies).
    Note: this is a NIH_MGC Library."

ORIGIN
  Query Match 100.0%; Score 19; DB 2; Length 877;
  Best Local Similarity 100.0%; Pred. No. 3.6e+02;
  Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCCTTTG 19
|||||
Db 225 CTTTACTTCATAGTCCTTTG 207

RESULT 31
BQ958341/c
LOCUS
DEFINITION
  BQ958341 899 bp mRNA linear EST 21-AUG-2002
  AGENCOURT_10037048 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6480481
  5', mRNA sequence.
ACCESSION
  BQ958341
VERSION
  BQ958341.1 GI:22373819
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)

```

```

ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
  Homnidae; Homo.
REFERENCE
  1 (bases 1 to 899)
  NIH-MGC http://mgs.nci.nih.gov/.
  National Institutes of Health, Mammalian Gene Collection (MGC)
  Unpublished (1999)
  Contact: Robert Strausberg, Ph.D.
  Email: cgabbs-r@mail.nih.gov
  Tissue Procurement: DCTD/DTF
  cDNA Library Preparation: Rubin Laboratory
  DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
  DNA Sequencing by: Agencourt Bioscience Corporation
  Clone distribution: MGC clone distribution information can be
  found through the I.M.A.G.E. Consortium/LLNL at:
  http://image.llnl.gov
  Plate: LLC2660 row: f column: 02
  High quality sequence stop: 636.
  Location/Qualifiers
    1..899
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone="IMAGE:6480481"
    /tissue_type="carcinoma, cell line"
    /lab_host="DH10B (phage-resistant)"
    /clone_lib="NIH MGC 40"
    /notes="Organ: prostate; Vector: pOTB7; Site 1: XhoI;
    Site 2: EcoRI; cDNA made by oligo-dt priming.
    Directionally cloned into EcoRI/XhoI sites using the
    following 5' adaptor: GGCACGAG(G). Library constructed by
    Ling Hong in the laboratory of Gerald M. Rubin (University
    of California, Berkeley) using ZAP-cDNA synthesis kit
    (Stratagene) and Superscript II RT (Life Technologies).
    Note: this is a NIH_MGC Library."

ORIGIN
  Query Match 100.0%; Score 19; DB 3; Length 899;
  Best Local Similarity 100.0%; Pred. No. 3.6e+02;
  Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCCTTTG 19
|||||
Db 587 CTTTACTTCATAGTCCTTTG 569

RESULT 32
BQ831564/c
LOCUS
DEFINITION
  BQ831564 901 bp mRNA linear EST 22-MAY-2001
  602765819F1 NIH_MGC_42 Homo sapiens cDNA clone IMAGE:4907885 5',
  mRNA sequence.
ACCESSION
  BQ831564
VERSION
  BQ831564.1 GI:14179151
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
  Homnidae; Homo.
REFERENCE
  1 (bases 1 to 901)
  NIH-MGC http://mgs.nci.nih.gov/.
  National Institutes of Health, Mammalian Gene Collection (MGC)
  Unpublished (1999)
  Contact: Robert Strausberg, Ph.D.
  Email: cgabbs-r@mail.nih.gov
  Tissue Procurement: ATCC
  cDNA Library Preparation: Ling Hong/Rubin Laboratory
  DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
  DNA Sequencing by: Incyte Genomics, Inc.
  Clone distribution: MGC clone distribution information can be
  found through the I.M.A.G.E. Consortium/LLNL at:
  http://image.llnl.gov

```

Plate: LLCMI813 row: e column: 06  
High quality sequence stop: 663.

# FEATURES

source Location/Qualifiers

1..901  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4907885"  
/tissue\_type="epithelioid carcinoma cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_42"  
/note="Organ: pancreas; Vector: pOTB7; Site 1: XhoI;  
Site 2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCAGGAG(G). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by Ling  
Hong in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library. |"

# ORIGIN

Query Match 100.0%; Score 19; DB 2; Length 901;  
Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGCTCTTG 19

Db 363 CTTTACTTCATAGCTCTTG 345

# RESULT 33

LOCUS BX372160 918 bp mRNA linear EST 27-APR-2004  
DEFINITION BX372160 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens  
cDNA clone CS0DK011VF17 3-PRIME, mRNA sequence.

ACCESSION BX372160

VERSION BX372160.1 GI:30452082

KEYWORDS EST.

SOURCE Homo sapiens (human)

# ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homniidae; Homo.

1 (bases 1 to 918)

Li, W.B., Gruber, C., Jessee, J. and Polayes, D.

Full-length cDNA libraries and normalization

Unpublished (2001)

Contact: Genoscope

Genoscope - Centre National de Sequencage

2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
end enriched, double-strand cDNA was digested with Not I and cloned  
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
was normalized. Library was constructed by Life Technologies, a  
division of Invitrogen. This sequence belongs to sequence cluster  
6601.f

For more information about this cluster, see  
http://www.genoscope.cns.fr/cdna?s=CS0BA1040ZE07\_CS03797\_1&c=6601.f

# FEATURES

source Location/Qualifiers

1..918  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DK011VF17"  
/cell\_type="HELA CELLS COT 25-NORMALIZED"  
/cell\_line="HELA"  
/clone\_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and cloned into the Not I and EcoR V

sites of the pCMVSPORT 6 vector. Library was normalized."

# ORIGIN

Query Match 100.0%; Score 19; DB 4; Length 918;  
Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGCTCTTG 19

Db 294 CTTTACTTCATAGCTCTTG 312

# RESULT 34

LOCUS BU543952/c

DEFINITION BU543952 923 bp mRNA linear EST 13-SEP-2002  
AGENCOURT\_10333041 NIH\_MGC\_40 Homo sapiens cDNA clone IMAGE:6576152  
5', mRNA sequence.

ACCESSION BU543952

VERSION BU543952.1 GI:22854435

KEYWORDS EST.

SOURCE Homo sapiens (human)

# ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homniidae; Homo.

1 (bases 1 to 923)

NIH-MGC http://mgc.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-rc@mail.nih.gov

Tissue Procurement: DCTD/DTF

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLCM2774 row: d column: 08

High quality sequence stop: 641.

# FEATURES

source Location/Qualifiers

1..923  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6576152"  
/tissue\_type="carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_40"  
/note="Organ: prostate; Vector: pOTB7; Site 1: XhoI;  
Site 2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCAGGAG(G). Library constructed by  
Ling Hong in the laboratory of Gerald M. Rubin (University  
of California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library."

# ORIGIN

Query Match 100.0%; Score 19; DB 3; Length 923;  
Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGCTCTTG 19

Db 587 CTTTACTTCATAGCTCTTG 569

# RESULT 35

LOCUS BQ685729/c

DEFINITION BQ685729

AGENCOURT\_8344399 NIH\_MGC\_110 Homo sapiens cDNA clone IMAGE:6251057

5', mRNA sequence.

940 bp mRNA linear EST 15-JUL-2002

```
ACCESSION BQ685729
VERSION BQ685729.1 GI:21811045
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 940)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2395 row: b column: 18
High quality sequence stop: 538.
FEATURES             Location/Qualifiers
     source          1..940
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CSODK011YF17"
                     /cell_type="HELA"
                     /clone_lib="HELA"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN
Query Match       100.0%; Score 19; DB 3; Length 940;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CTTTACTTCATAGTCCTTTG 19
|||||
Db 728 CTTTACTTCATAGTCCTTTG 710
|||||

RESULT 36
LOCUS BQ67433/c
DEFINITION BQ67433 1023 bp mRNA linear EST 02-APR-2002
AGENCOURT_6758944 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5754944
5', mRNA sequence.
ACCESSION BQ67433
VERSION BQ67433.1 GI:19896479
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 1023)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12793 row: c column: 09
High quality sequence stop: 649.
FEATURES             Location/Qualifiers
     source          1..1023
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5754944"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_115"
                     /note="Organ: pooled brain, lung, testis; Vector:
                     pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA
                     source anonymous pool of 6 male brains, age range 23-27; 1
                     male lung, age 27; and 1 male testis, age 69. Library is
                     oligo-dT primed and directionally cloned (EcoRV site is
```

destroyed upon cloning). Average insert size 1.8 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 021. Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 100.0%; Score 19; DB 3; Length 1023;  
Best Local Similarity 100.0%; Pred. No. 3.7e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19  
|||||  
Db 444 CTTTACTTCATAGTCTTTG 426

## RESULT 38

CR597125/c  
LOCUS 1830 bp mRNA linear HTC 21-JUL-2004  
DEFINITION full-length cDNA clone CSODK011YF17 of HeLa cells Cot 25-normalized of Homo sapiens (human).

ACCESSION CR597125.1 GI:50477932

VERSION HTC; CNSLT\_CDNA.

KEYWORDS Homo sapiens (human)

SOURCE Homo sapiens

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 1830)

AUTHORS Li, W.B., Gruber, C., Jesse, J. and Polayes, D.

TITLE Full-length cDNA libraries and normalization

JOURNAL Unpublished

REMARK Contact: Feng Liang Email: fliang@lifetech.com URL: http://fulllength.invitrogen.com/Invitrogen Corporation 1600

Paradise Avenue

REFERENCE 2 (bases 1 to 1830)

Genoscope.

TITLE Direct Submission

JOURNAL Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :

BP 191 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr)  
- Web : www.genoscope.cns.fr

COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life technologies, a division of invitrogen.

## FEATURES

source  
Location/Qualifiers  
1..1830  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clones="CSODK011YF17"  
/tissue\_type="HeLa cells Cot 25-normalized"  
/plasmid="pCMVSPORT\_6"

## ORIGIN

Query Match 100.0%; Score 19; DB 6; Length 1830;  
Best Local Similarity 100.0%; Pred. No. 3.9e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19  
|||||  
Db 1159 CTTTACTTCATAGTCTTTG 1141

## RESULT 39

DU976902  
LOCUS 490 bp DNA linear GSS 03-JAN-2006  
DEFINITION AUAC-aaa07d08.g1 Ascaris suum whole genome shotgun library (PMAJ\_4 GSS) Ascaris suum genomic, genomic survey sequence.

ACCESSION DU976902

VERSION DU976902.1 GI:84308300

## KEYWORDS

## SOURCE

## ORGANISM

Ascaris suum (pig roundworm)  
Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida;  
Ascaridoidea; Ascarididae; Ascaris.

## REFERENCE

## AUTHORS

1 (bases 1 to 490)

Mitreva, M., McCarter, J.P., Ronko, I., Martin, J., Wylie, T., Dante, M.,

Waterston, R.H., Clifton, S.W. and Wilson, R.

Genome Survey sequences from the parasitic nematode Ascaris suum

Unpublished (2005)

JOURNAL

COMMENT

Contact: Mitreva M

Washington University in St. Louis

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: nematode@watson.wustl.edu

Genomic DNA provided by Richard E. Davis at the University of

Colorado medical school. DNA sequenced by Washington University

Genome Sequencing Center

Class: shotgun.

Location/Qualifiers

1..490

/organism="Ascaris suum"

/mol\_type="genomic DNA"

/db\_xref="taxon:6253"

/dev\_stage="32-64 cell embryos"

/lab\_host="GS10"

/clone\_lib="Ascaris suum whole genome shotgun library

(PMAJ\_4 GSS)"

/note="Vector: pOTW13; Site 1: BstX1; Site 2: BstX1;

Ascaris suum genomic DNA was randomly sheared,

end-repaired and size fractionated to enrich for 6-6.2 kb

fragments. Genomic DNA was provided by Richard E. Davis

at the University of Colorado medical school. Sequencing

by Washington University Genome Sequencing Center, St.

Louis, MO."

Query Match 94.7%; Score 18; DB 14; Length 490;

Best Local Similarity 100.0%; Pred. No. 1e+03;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TTTACTTCATAGTCTTTG 19

|||||

Db 81 TTTACTTCATAGTCTTTG 98

|||||

RESULT 40

DU069153/c

LOCUS 811 bp DNA linear GSS 12-AUG-2005

DEFINITION 139562 Tomato HindIII BAC Library Lycopersicon esculentum genomic

clone LE HBa0107M06 5, genomic survey sequence.

ACCESSION DU069153

VERSION DU069153.1 GI:72505294

KEYWORDS GSS.

SOURCE Lycopersicon esculentum (Solanum lycopersicum)

ORGANISM Lycopersicon esculentum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;

asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.

REFERENCE 1 (bases 1 to 811)

Mueller, L.A., Buels, R.M., Wang, Y., Tanksley, S.D., Giovannoni, J.J.,

Van Eck, J. and Stack, S.

BAC end sequencing from three Solanum lycopersicon libraries

Unpublished (2005)

Other GSSs: 316712

Contact: Lukas Mueller

Tanksley Lab, Dept. of Plant Breeding

Cornell University

251 Emerson Hall, Ithaca, NY 14853, USA

Tel: 607-255-6557

Fax: 607-255-6683

## GSS.

Ascaris suum (pig roundworm)

Ascaris suum

Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida;

Ascaridoidea; Ascarididae; Ascaris.

REFERENCE

AUTHORS

1 (bases 1 to 490)

Mitreva, M., McCarter, J.P., Ronko, I., Martin, J., Wylie, T., Dante, M.,

Waterston, R.H., Clifton, S.W. and Wilson, R.

Genome Survey sequences from the parasitic nematode Ascaris suum

Unpublished (2005)

JOURNAL

COMMENT

Contact: Mitreva M

Washington University in St. Louis

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: nematode@watson.wustl.edu

Genomic DNA provided by Richard E. Davis at the University of

Colorado medical school. DNA sequenced by Washington University

Genome Sequencing Center

Class: shotgun.

Location/Qualifiers

1..490

/organism="Ascaris suum"

/mol\_type="genomic DNA"

/db\_xref="taxon:6253"

/dev\_stage="32-64 cell embryos"

/lab\_host="GS10"

/clone\_lib="Ascaris suum whole genome shotgun library

(PMAJ\_4 GSS)"

/note="Vector: pOTW13; Site 1: BstX1; Site 2: BstX1;

Ascaris suum genomic DNA was randomly sheared,

end-repaired and size fractionated to enrich for 6-6.2 kb

fragments. Genomic DNA was provided by Richard E. Davis

at the University of Colorado medical school. Sequencing

by Washington University Genome Sequencing Center, St.

Louis, MO."

Query Match 94.7%; Score 18; DB 14; Length 490;

Best Local Similarity 100.0%; Pred. No. 1e+03;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TTTACTTCATAGTCTTTG 19

|||||

Db 81 TTTACTTCATAGTCTTTG 98

|||||

Email: sgn-feedback@sgn.cornell.edu  
 Insert Length: 94875 Std Error: 0.00  
 Plate: 107 row: M column: 6  
 Seq primer: T7  
 Class: BAC ends  
 High quality sequence start: 44  
 High quality sequence stop: 594.  
 Location/Qualifiers  
 1. .811  
 /organism="Lycopersicon esculentum"  
 /mol\_type="genomic DNA"  
 /cultivar="Heinz 1706"  
 /db\_xref="taxon:4081"  
 /clone="LE\_HBa0107M06"  
 /lab\_host="E. coli"  
 /clone\_lib="Tomato HindIII BAC Library"  
 /note="Vector: pBelOBAC11; Site\_1: HindIII"

ORIGIN  
 Query Match 94.7%; Score 18; DB 13; Length 811;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+03;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TTTACTTCATAGCTCTTTG 19  
 |||||  
 Db 765 TTTACTTCATAGCTCTTTG 748

RESULT 41  
 CF780868/c  
 LOCUS  
 DEFINITION AGENCOURT\_15939245 NIH\_MGC\_219 Homo sapiens cDNA clone  
 IMAGE:30523569 5', mRNA sequence.  
 CF780868  
 VERSION CF780868.1 GI:37740645  
 EST.  
 KEYWORDS Homo sapiens (human)  
 SOURCE  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 916)  
 NIH-MGC http://mgs.nci.nih.gov/.  
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
 TITLE Unpublished (1999)  
 JOURNAL  
 COMMENT Contact: Daniela S. Gerhard, Ph.D.  
 Office of Cancer Genomics  
 National Cancer Institute / NIH  
 Bldg. 31 Rm10A07 Bethesda, MD 20892  
 Email: cgabbs-r@mail.nih.gov  
 Tissue Procurement: James Martin, University of Iowa  
 cDNA Library Preparation: M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: NDAM601 row: 1 column: 10  
 High quality sequence stop: 651.  
 Location/Qualifiers  
 1. .916  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:30523569"  
 /tissue\_type="Pooled Chondrosarcoma Tumor cells"  
 /lab\_host="DHI08 (T1 phage-resistant)"  
 /clone\_lib="NIH\_MGC\_219"  
 /notes="Vector: pYX-Asc; Site\_1: EcoRI; Site\_2: NotI;  
 Library is oligo-dT primed and directionally cloned  
 Denatured RNA was size fractionated on a 1% agarose gel.  
 First strand cDNA synthesis was primed with oligo-dT  
 primer containing a Not I site. Double strand cDNA was

size selected according to mRNA size fraction, ligated with  
 EcoR I adaptor, digested with Not I and then cloned  
 directionally into pYX-Asc vector. Average insert size  
 0.5-1kb. Adaptors 5' (AATTCGACGAGG)3' and 5'd  
 (CCTCGTGGCG)3'. 3' Linker sequence - GCGGCGCTGAGAGCC T18.  
 Sequencing primers 3' end: T3 promoter primer 5'd  
 (ATTACCCCTCACTAAGGGA)3'. 5' End: T7 promoter primer 5'd  
 (TAATACGACTCACTATAGG)3'. Library was constructed in the  
 laboratory of M. Bento Soares. Average insert size 2-3kb.  
 Note: this is a NIH\_MGC Library."

ORIGIN  
 Query Match 94.7%; Score 18; DB 5; Length 916;  
 Best Local Similarity 94.7%; Pred. No. 1.1e+03;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGCTCTTTG 19  
 |||||  
 Db 708 CTTTACTTCATAGCTCTTTG 690

RESULT 42  
 BI193620/c  
 LOCUS  
 DEFINITION BI193620 1125 bp mRNA linear EST 10-JUL-2001  
 602946519F1 NIH\_MGC\_42 Homo sapiens cDNA clone IMAGE:5089900 5',  
 mRNA sequence.  
 BI193620  
 VERSION BI193620.1 GI:14648640  
 EST.  
 KEYWORDS Homo sapiens (human)  
 SOURCE  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 1125)  
 NIH-MGC http://mgs.nci.nih.gov/.  
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
 TITLE Unpublished (1999)  
 JOURNAL  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-r@mail.nih.gov  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLCM1851 row: e column: 05  
 High quality sequence stop: 529.  
 Location/Qualifiers  
 1. .1125  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5089900"  
 /tissue\_type="epithelioid carcinoma cell line"  
 /lab\_host="DHI08 (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_42"  
 /notes="Organ: pancreas; Vector: pOTB7; Site 1: XhoI;  
 Site 2: EcoRI; cDNA made by oligo-dT priming.  
 Directionally cloned into EcoRI/XhoI sites using the  
 following 5' adaptor: GGCACGAG(G). Size-selected >500bp  
 for average insert size 1.8kb. Library constructed by Ling  
 Hong in the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies).  
 Note: this is a NIH\_MGC Library."

ORIGIN  
 Query Match 94.7%; Score 18; DB 2; Length 1125;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+03;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



```

Qy 2 TTTACTTCATAGTCTTTG 19
Db 334 TTTACTTCATAGTCTTTG 317

RESULT 43
DE194703
LOCUS
DEFINITION Branchiostoma floridae DNA, clone: CH302-076_I05_T7, genomic survey
sequence.
ACCESSION DE194703
VERSION DE194703.1 GI:73603031
KEYWORDS GSS.
SOURCE Branchiostoma floridae (Florida lancelet)
ORGANISM Branchiostoma floridae
Eukaryota: Metazoa; Chordata; Cephalochordata; Branchiostomidae;
Branchiostoma.
REFERENCE 1
AUTHORS Fujiyama,A., Toyoda,A., Kuroki,Y. and Sakaki,Y.
TITLE BAC end sequences of CHORI-302 Amphioxus Library
JOURNAL Published Only in Database (2005)
REFERENCE 2 (bases 1 to 400)
AUTHORS Fujiyama,A.
TITLE Direct Submission
JOURNAL Submitted (12-AUG-2005) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, 230-0045, Japan
(E-mail:afujiyam@gsc.riken.jp, URL:http://stt.gsc.riken.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
PRIMERS
Sequencing : T7
LIBRARY
Vector : PTARBAC2.1
R.Site 1 : EcoRI
R.Site 2 : EcoRI.
Location/Qualifiers
1. 400
/organism="Branchiostoma floridae"
/mol_type="genomic DNA"
/db_xref="taxon:7739"
/clone="CH302-076_I05_T7"
/sex="male"
/tissue_type="sperm"
/clone_lib="CHORI0302 Amphioxus genomic BAC library"
/notes="Common name:amphioxus"

ORIGIN
Query Match 91.6%; Score 17.4; DB 14; Length 400;
Best Local Similarity 94.7%; Pred. No. 1.9e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19
Db 198 CTTTATTTCATAGTCTTTG 216

RESULT 44
AQ435685/c
LOCUS
DEFINITION HS_5148_B1_F10_SP6E RPCI-11 Human Male BAC Library Homo sapiens
genomic clone Plate=724 Col=19 Row=L, genomic survey sequence.
ACCESSION AQ435685
VERSION AQ435685.1 GI:4547024
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homniidae; Homo.
REFERENCE 1 (bases 1 to 429)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.

Qy 2 TTTACTTCATAGTCTTTG 19
Db 334 TTTACTTCATAGTCTTTG 317

TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
PUBMED 10449764
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Resear h Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 724 row: L column: 19
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 429.
FEATURES
Location/Qualifiers
1. 429
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="plate=724 Col=19 Row=L"
/sex="male"
/clone_lib="RPCI-11 Human Male BAC Library"
/notes="Vector: pBACe3.6; Site 1: EcoRI; Site 2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACe3.6 vector at EcoRI sites"

ORIGIN
Query Match 91.6%; Score 17.4; DB 11; Length 429;
Best Local Similarity 94.7%; Pred. No. 1.9e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19
Db 176 CTTTATTTCATAGTCTTTG 158

RESULT 45
AW664308/c
LOCUS
DEFINITION hi09e10.x1 NCI_CGAP_GUI Homo sapiens cDNA clone IMAGE:2971818 3',
mRNA sequence.
ACCESSION AW664308
VERSION AW664308.1 GI:7456849
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homniidae; Homo.
REFERENCE 1 (bases 1 to 498)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapsb@emall.nih.gov
Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life
Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The
I.M.A.G.E. Consortium DNA Sequencing by: Washington University
Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
image.llnl.gov/image/html/iresources.shtml

```

Query match	91.8%	Score 17.4	DB 8	Length 523
Best Local Similarity	94.7%	Pred. No. 2e+03		
Matches 18; Conservative	0	Mismatches 1	Indels 0	Gaps 0
LOCUS	DT490210	565 bp	mRNA	linear
DEFINITION	WS02544.B21.F11	Populus trichocarpa	cDNA clone	
EST 29-AUG-2005				



Query Match 91.6%; Score 17.4; DB 1; Length 603;  
 Best Local Similarity 94.7%; Pred. No. 2e+03; 1; Indels 0; Gaps 0;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCTTTG 19  
 ||||| ||||| ||||| |||||  
 Db 109 CTTTACTTCATAGTCTTTG 91

RESULT 51  
 AZ353695/c  
 LOCUS  
 DEFINITION 630 bp DNA linear GSS 02-OCT-2000  
 clone UUGC1M0092L09 R, genomic survey sequence.  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Mus musculus (house mouse)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridea; Muridae; Murinae; Mus.  
 1 (bases 1 to 630)  
 Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,  
 Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von  
 Niederhausern,A. and Wright,D. Weiss,R.  
 Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts  
 Unpublished (2000)  
 Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 34112, USA  
 Rn. 308 Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plates: 0092 row: L column: 09  
 Seq primer: CACACAGGAACAGCTATGACC  
 Class: plasmid ends  
 High quality sequence stop: 630.

FEATURES  
 source  
 1..630  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGC1M0092L09"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, TI-resistant, F-"  
 /clone\_lib="Mouse 10kb plasmid UUGC1M library"  
 /notes="Vector: FWD42nv; Purified genomic DNA from M.  
 musculus C57BL/6J (male) was obtained from the Jackson  
 Laboratory Mouse DNA Resource  
 (http://www.jax.org/resources/documents/dnares/). The DNA  
 was hydrodynamically sheared by repeated passage through a  
 0.005 inch orifice at constant velocity. The sheared DNA  
 was blunt end-repaired with T4 DNA polymerase and T4  
 polynucleotide kinase. Adaptor oligonucleotides were  
 ligated to the blunt ends in high molar excess. The  
 adaptor DNA was purified and size-selected for a 9.5 to  
 10.5 kb range using preparative agarose gel  
 electrophoresis. Vector DNA was prepared from a derivative  
 of pWD42 [Gill4732114|gb|AF129072.1], a copy-number  
 inducible derivative of plasmid R1. The vector was ligated  
 with adaptors complementary to the insert adaptors and  
 purified. The sheared, adaptor mouse DNA was annealed to  
 adaptor vector DNA, and transformed into  
 chemically-competent E. coli XL10-Gold (Stratagene) cells  
 and selected for ampicillin resistance."

## ORIGIN

Query Match 91.6%; Score 17.4; DB 11; Length 630;  
 Best Local Similarity 94.7%; Pred. No. 2e+03; 1; Indels 0; Gaps 0;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCTTTG 19  
 ||||| ||||| ||||| |||||  
 Db 77 CTTTAAATTCATAGTCTTTG 59

RESULT 52  
 AG158060  
 LOCUS  
 DEFINITION 635 bp DNA linear GSS 09-JAN-2002  
 Pan troglodytes DNA, clone: RP43-023C18.TJ, genomic survey  
 sequence.  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Pan troglodytes (chimpanzee)  
 Pan troglodytes  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homnidae; Pan.

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 1  
 Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,  
 Totoki,Y., Watanabe,H. and Sakaki,Y.  
 BAC end sequences of Library RPCI-43  
 Unpublished  
 2 (bases 1 to 635)  
 Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,  
 Totoki,Y., Watanabe,H. and Sakaki,Y.  
 Direct Submission  
 Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical  
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
 1-7-22 Suehiro-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 (E-mail:chimpsbgsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,  
 Tel:81-45-503-9111, Fax:81-45-503-9170)  
 Clones are derived from the chimpanzee BAC library RPCI-43 This BAC  
 end was generated during the R&D process and may have higher chance  
 of clone tracking errors.

PRIMERS  
 Sequencing: TJ  
 LIBRARY  
 Vector : pBACE3.6  
 R.Site 1 : EcoRI  
 R.Site 2 : EcoRI  
 Location/Qualifiers  
 1..635  
 /organism="Pan troglodytes"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9598"  
 /clone="RP43-023C18.TJ"  
 /sex="male"  
 /cell\_type="lymphocytes"  
 /clone\_lib="RPCI-43 Chimpanzee Male BAC Library"

## COMMENT

FEATURES  
 source

1..635  
 /organism="Pan troglodytes"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9598"  
 /clone="RP43-023C18.TJ"  
 /sex="male"  
 /cell\_type="lymphocytes"  
 /clone\_lib="RPCI-43 Chimpanzee Male BAC Library"

## ORIGIN

Query Match 91.6%; Score 17.4; DB 14; Length 635;  
 Best Local Similarity 94.7%; Pred. No. 2e+03; 1; Indels 0; Gaps 0;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCTTTG 19  
 ||||| ||||| ||||| |||||  
 Db 482 CTTTACTTCATAGTCTTTG 500

RESULT 53  
 AZ840748  
 LOCUS  
 DEFINITION 644 bp DNA linear GSS 20-FEB-2001  
 2M0138K24F Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
 clone UUGC2M0138K24 F, genomic survey sequence.  
 AZ840748

VERSION A2840748.1 GI:13010656  
 KEYWORDS GSS.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridea; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 644)  
 AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mamoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhauser, A. and Wright, D., Weiss, R.  
 TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Robert B. Weiss  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: dunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0138 row: K column: 24  
 Seq primer: CTTGTAAACGACGCCAGT  
 Class: plasmid ends  
 High quality sequence stop: 644.  
 Location/Qualifiers  
 1. 644  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGC2M0138K24"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
 /clone\_lib="Mouse 10kb plasmid UGCIM library"  
 /notes="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (gi4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

# FEATURES

source  
 1. 657  
 /organism="Euphorbia esula"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:3993"  
 /clone="CV03074B2E03.f1"  
 /lab\_host="DH10B"  
 /clone\_lib="CV03-normalized library"  
 /notes="Vector: pBS II SK(+); Site 1: EcoRI(5' side of insert); Site 2: NotI (3' side of insert); Tissue: Seasonal crown buds from outdoor plants, whole plant tissue (leaf, stem, flowers, seeds, shoots, galls, root and crown buds) from outdoor plants, cold acclimated tissues (leaf, stem, and crown buds), crown and root buds from decapitated greenhouse p ants, and seedlings. The library was constructed and normalized as described by Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996), Genome Research 6(9): 791-806. An identifying tag was added at the 3' during cDNA synthesis: insertAAAAAAAAAAAAAAAAATGCGT."

# ORIGIN

Query Match 91.6%; Score 17.4; DB 11; Length 644;  
 Best Local Similarity 94.7%; Pred. No. 2e+03;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGCTCTTTG 19  
 |||||  
 Db 156 CTTTCTTCATAGCTCTTTG 174

# RESULT 54

DV148009  
 LOCUS 657 bp mRNA linear EST 03-OCT-2005  
 DEFINITION CV03074B2E03.f1 CV03-normalized library Euphorbia esula cDNA clone  
 CV03074B2E03.f1 5, mRNA sequence.

# ACCESSION VERSION KEYWORDS SOURCE ORGANISM

DV148009  
 DV148009.1 GI:76857016  
 EST.  
 Euphorbia esula (leafy spurge)  
 Euphorbia esula  
 Eukaryota; Viridiplantae; Streptophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids I; Malpighiales; Euphorbiaceae; Euphorbioideae; Euphorbiaceae; Euphorbia.  
 REFERENCE 1 (bases 1 to 657)  
 AUTHORS Anderson, J.V., Horvath, D.P., Thimmapuram, J., Liu, L., Hernandez, A., Kim, W.R. and Mikel, M.  
 TITLE Direct submission, Anderson, J.V. 2005  
 JOURNAL Unpublished (2005)  
 COMMENT Contact: James V. Anderson  
 USDA/ARS  
 1605 Albrecht Blvd, Biosciences Research Laboratory, Fargo, ND 58105, USA  
 Tel: 701 239 1263  
 Fax: 701 239 1252  
 Email: andersjv@fargo.ars.usda.gov  
 Adaptors and tags in 5'-end sequenced clones: (Vector) . . . TAAGCTTGATTCG (End Vector) (Start EcoRI adaptor) AATTCATTTGTTGG (End EcoRI adaptor) (Start Insert) . . . AAAAAAAAAAAAAA (End Insert) (Start Tag) TCGGT (End Tag) (Start NotI site/Vector) CGCGCCGCCACCGCGG . . . Base Calling/Quality Scores: PHRED from Washington University Genome Center. Vector Trimming: Cross-match from Washington University Genome Center PHRAP suite. Low quality bases (phred score < 20) were trimmed from both ends of the sequence by an in-house script. This sequence is vector free and at least 200 bp in length.  
 PCR Primers  
 FORWARD: TAATACGACTCACTATAGG (T7)  
 BACKWARD: ATTAACCCCTCACTAAG (T3)  
 Insert Length: 657 Std Error: 0.00  
 Plate: CV03074B2 row: E column: 03  
 Seq primer: TAATACGACTCACTATAGG (T7)  
 High quality sequence stop: 657.  
 Location/Qualifiers  
 1. 657  
 /organism="Euphorbia esula"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:3993"  
 /clone="CV03074B2E03.f1"  
 /lab\_host="DH10B"  
 /clone\_lib="CV03-normalized library"  
 /notes="Vector: pBS II SK(+); Site 1: EcoRI(5' side of insert); Site 2: NotI (3' side of insert); Tissue: Seasonal crown buds from outdoor plants, whole plant tissue (leaf, stem, flowers, seeds, shoots, galls, root and crown buds) from outdoor plants, cold acclimated tissues (leaf, stem, and crown buds), crown and root buds from decapitated greenhouse p ants, and seedlings. The library was constructed and normalized as described by Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996), Genome Research 6(9): 791-806. An identifying tag was added at the 3' during cDNA synthesis: insertAAAAAAAAAAAAAAAAATGCGT."

# FEATURES

source  
 1. 657  
 /organism="Euphorbia esula"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:3993"  
 /clone="CV03074B2E03.f1"  
 /lab\_host="DH10B"  
 /clone\_lib="CV03-normalized library"  
 /notes="Vector: pBS II SK(+); Site 1: EcoRI(5' side of insert); Site 2: NotI (3' side of insert); Tissue: Seasonal crown buds from outdoor plants, whole plant tissue (leaf, stem, flowers, seeds, shoots, galls, root and crown buds) from outdoor plants, cold acclimated tissues (leaf, stem, and crown buds), crown and root buds from decapitated greenhouse p ants, and seedlings. The library was constructed and normalized as described by Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996), Genome Research 6(9): 791-806. An identifying tag was added at the 3' during cDNA synthesis: insertAAAAAAAAAAAAAAAAATGCGT."

# ORIGIN

Query Match 91.6%; Score 17.4; DB 10; Length 657;  
 Best Local Similarity 94.7%; Pred. No. 2e+03;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGCTCTTTG 19  
 |||||  
 Db 113 CTTTACTTCATATCTTTG 131

# RESULT 55

DT482123/c  
 LOCUS 703 bp mRNA linear EST 29-AUG-2005  
 DEFINITION WS02533\_BR\_P14 PT-MB-N-A-15 Populus trichocarpa cDNA clone  
 WS02533\_P14 5, mRNA sequence.



```

PUBMED 15353603
COMMENT Other ESTs: M104E06, M104E06.SP8
        Contact: Bo Segerman
        Umea Plant Science Center, Department of Plant Physiology
        Umea University
        901 87 Umea, Sweden
        Tel: +46 90 786 5279
        Fax: +46 90 786 6676
        Email: bo.segerman@plantphys.umu.se.

FEATURES
    source
        1..735
        /organism="Populus trichocarpa"
        /mol_type="mRNA"
        /db_xref="taxon:3694"
        /clone="M104E06"
        /tissue_type="Female catkins"
        /clone_lib="Populus female catkins cDNA library"

ORIGIN
    Query Match      91.6%; Score 17.4; DB 9; Length 735;
    Best Local Similarity 94.7%; Pred. No. 2e+03;
    Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19
    ||||| ||||| |||||
Db 83 CTTTACTACATAGTCTTTG 65

RESULT 58
DV140694
LOCUS DV140694 740 bp mRNA linear EST 03-OCT-2005
DEFINITION CV03126A2C04.f1 CV03-normalized library Euphorbia esula cDNA clone
ACCESSION DV140694
VERSION DV140694.1 GI:76849701
KEYWORDS EST.
SOURCE Euphorbia esula (leafy spurge)
ORGANISM Euphorbia esula
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
          rosids; eurosids I; Malpighiales; Euphorbiaceae; Euphorbioideae;
          Euphorbiaceae; Euphorbia.
REFERENCE 1 (bases 1 to 740)
AUTHORS Anderson,J.V., Horvath,D.P., Thimmapuram,J., Liu,L., Hernandez,A.,
          Kim,W.R. and Mikel,M.
TITLE Direct submission, Anderson,J.V. 2005
JOURNAL Unpublished (2005)
COMMENT Contact: James V. Anderson
          USDA/ARS
          1605 Albrecht Blvd, Biosciences Research Laboratory, Fargo, ND
          58105, USA
          Tel: 701 239 1263
          Fax: 701 239 1252
          Email: andersjv@fargo.ars.usda.gov
          Adaptors and tags in 5'-end sequenced clones:
          (Vector) TAGCTTGATATCG (End Vector) (Start
          EcoRI adaptor) AATTCATGTGTTGG (End EcoRI adaptor) (Start
          Insert) . . . AAAAAAAAAAAAAA (End Insert) (Start tag) TGCCT (End
          Tag) (Start NotI site/Vector) GCGCGCCACCGCGG. . . Base
          Calling/Quality Scores: PHRED from Washington University Genome
          Center. Vector Trimming: Cross match from Washington University
          Genome Center PHRAP suite. Low quality bases (Phred score < 20)
          were trimmed from both ends of the sequence by an in-house script.
          This sequence is vector free and at least 200 bp in length.
          PCR Primers
          FORWARD: TAATACGACTCACTATAGG (T7)
          BACKWARD: ATTAACCTCACTAAAG (T3)
          Insert Length: 740 Std Error: 0.00
          Plate: CV03126A2 row: C column: 04
          Seq primer: TAATACGACTCACTATAGG (T7)
          High quality sequence stop: 740.
          Location/Qualifiers
          1..740

```

```

/organism="Euphorbia esula"
/mol_type="mRNA"
/db_xref="taxon:3993"
/clone="CV03126A2C04.f1"
/lab_host="DH10B"
/clone_lib="CV03-normalized library"
/notes="vector: pBS II SK(+); Site 1: EcoRI(5' side of
insert); Site 2: NotI (3' side of insert); Tissue:
Seasonal crown buds from outdoor plants, whole plant
tissue (leaf, stem, flowers, seeds, shoots, galls, root
and crown buds) from outdoor plants, cold acclimated
tissues (leaf, stem, and crown buds), crown and root buds
from decapitated greenhouse p ants, and seedlings. The
library was constructed and normalized as described by
Bonaldi, M.F., Lennon, G. and Soares, M.B. (1996), Genome
Research 6(9): 791-806. An identifying tag was added at the
3' during cDNA synthesis: insertAAAAAAAAAAAAAAAAATCGCT."

ORIGIN
    Query Match      91.6%; Score 17.4; DB 10; Length 740;
    Best Local Similarity 94.7%; Pred. No. 2e+03;
    Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19
    ||||| ||||| |||||
Db 20 CTTTACTTCATATCTTTG 38

RESULT 59
CO081056/c
LOCUS GR_Ea44N24_r GR_Ea Gossypium raimondii cDNA clone linear EST 15-JUN-2004
DEFINITION GR_Ea44N24_r GR_Ea Gossypium raimondii cDNA clone GR_Ea44N24 3',
          mRNA sequence.
ACCESSION CO081056
VERSION CO081056.1 GI:48750537
KEYWORDS EST.
SOURCE Gossypium raimondii
ORGANISM Gossypium raimondii
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
          rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
REFERENCE 1 (bases 1 to 765)
AUTHORS Kim,H., Yu,X., Kudrna,D., Hatfield,J., Stum,D., Mueller,C.,
          Udall,J.A., Rapp,R.A., Wendel,J.F., Rao,K., Soderlund,C. and
          Wing,R.A.
TITLE Global assembly of Cotton ESTs
JOURNAL Unpublished (2004)
COMMENT Contact: Rod A. Wing
          Arizona Genomics Institute
          The University of Arizona
          Forbes Building Room 303, Tucson, AZ, 85721-0036, USA
          Tel: 520 626 9595
          Fax: 520 621 1259
          Email: rwing@genome.arizona.edu
          Plate: 44 row: N column: 24.
          Location/Qualifiers
          1..765
          /organism="Gossypium raimondii"
          /mol_type="mRNA"
          /db_xref="taxon:29730"
          /clone="GR_Ea44N24"
          /tissue_type="whole seedlings"
          /dev_stage="first true leaves"
          /lab_host="DH10B"
          /clone_lib="GR_Ea"
          /notes="vector: pCMV.SPORT-6.1; Site 1: NotI; Site 2:
          EcoRV; Library made by invitrogen with RNA supplied by
          Wendel lab. Directional cloned into NotI-EV. Colonies
          plated/picked by AGI. More glycerol clones held in -80."

ORIGIN
    Query Match      91.6%; Score 17.4; DB 8; Length 765;
    Best Local Similarity 94.7%; Pred. No. 2e+03;

```

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCTTTG 19  
 ||||| ||||| |||||  
 Db 431 CTTTACTTCATAGTCTTTG 413

RESULT 60  
 CV276848/c 766 bp mRNA linear EST 22-SEP-2004  
 LOCUS  
 DEFINITION  
 WS0141\_B21\_P08 PTX-D-IL-A-5 Populus trichocarpa x Populus deltoides  
 cDNA clone WS0141\_P08 3', mRNA sequence.

ACCESSION  
 CV276848  
 VERSION  
 CV276848.1 GI:52529823  
 KEYWORDS  
 EST.

SOURCE  
 Populus trichocarpa x Populus deltoides  
 ORGANISM  
 Populus trichocarpa x Populus deltoides  
 Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
 rosids; eurosids I; Malpighiales; Salicaceae; Populus.  
 1 (bases 1 to 766)

REFERENCE  
 AUTHORS  
 Ralph,S., Cooper,D., Kolosova,N., Oddy,C., Butterfield,Y.,  
 Kirkpatrick,R., Liu,J., Palmquist,D., Scott,J., Barber,S., Yang,G.,  
 Babakaiff,R., Brown-John,M., Chand,S., Featherstone,R., Masson,A.,  
 Mayo,M., Moran,J., Olson,T., Wong,D., Ritland,C.E., Siddiqui,A.,  
 Holt,R., Jones,S., Marra,M., Ellis,B.E., Douglas,C., Ritland,K. and  
 Bohlmann,J.

TITLE  
 The poplar transcriptome: Analysis of expressed sequence tags from  
 multiple cDNA libraries

JOURNAL  
 COMMENT  
 Contact: Joerg Bohlmann  
 Genome BC forest genomics program  
 University of British Columbia  
 Michael Smith Laboratories, 6174 University Boulevard, Rm. 237,  
 Vancouver, British Columbia, Canada, V6T 1Z3  
 Tel: 1-604-822-0282  
 Fax: 1-604-822-2114  
 Email: bohlmann@msl.ubc.ca  
 Plate: WS0141 row: P column: 08  
 High quality sequence stop: 766  
 POLYA=yes.

FEATURES  
 Location/Qualifiers  
 1..766  
 /organism="Populus trichocarpa x Populus deltoides"  
 /mol\_type="mRNA"  
 /cultivar="H11-11"  
 /db\_xref="taxon:3695"  
 /clone="WS0141\_P08"  
 /sex="Male"  
 /lab\_host="E. coli DH10B T1 phage resistant cells"  
 /clone\_lib="PTX-D-IL-A-5"  
 /note="Vector: pBluescript II SK (+) XR; Site 1: EcoRI (5'  
 end of cDNA); Site 2: XhoI (3' end of cDNA); Sapling trees  
 two metres in height and grown under greenhouse conditions  
 were exposed to continuous feeding by Malacosoma disstria  
 Hubner (forest tent caterpillar) mid-instar larvae caged  
 on the sapling using mesh bags. Mature leaves from within  
 the caged region were collected 2 hours, 12 hours, 24  
 hours and 48 hours after the onset of treatment. mRNA was  
 isolated from each tissue source independently and equal  
 quantities of mRNA from each tissue were then pooled. cDNA  
 was prepared from 5 micrograms of mRNA and vectorially  
 ligated into the pBluescript II SK (+) XR vector using the  
 pBluescript II XR cDNA Library Construction Kit according  
 to manufacturer's instructions with modifications  
 (Stratagene). Plasmid DNA was then transformed by  
 electroporation into DH10B cells (Invitrogen) for  
 propagation."

Query Match 91.6%; Score 17.4; DB 8; Length 766;  
 Best Local Similarity 94.7%; Pred. No. 2e+03;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCTTTG 19  
 ||||| ||||| |||||  
 Db 431 CTTTACTTCATAGTCTTTG 413

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCTTTG 19  
 ||||| ||||| |||||  
 Db 50 CTTTACTTCATAGTCTTTG 32

RESULT 61  
 DT478689 779 bp mRNA linear EST 29-AUG-2005  
 LOCUS  
 DEFINITION  
 WS02523\_BR\_I07 PT-MB-N-A-15 Populus trichocarpa cDNA clone  
 WS02523\_I07 5', mRNA sequence.

ACCESSION  
 DT478689  
 VERSION  
 DT478689.1 GI:73875951  
 KEYWORDS  
 EST.

SOURCE  
 Populus trichocarpa (Populus balsamifera subsp. trichocarpa)  
 ORGANISM  
 Populus trichocarpa  
 Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
 rosids; eurosids I; Malpighiales; Salicaceae; Populus.  
 1 (bases 1 to 779)

REFERENCE  
 AUTHORS  
 Ralph,S., Cooper,D., Kolosova,N., Oddy,C., Butterfield,Y.,  
 Kirkpatrick,R., Liu,J., Palmquist,D., Scott,J., Barber,S., Yang,G.,  
 Babakaiff,R., Brown-John,M., Chand,S., Featherstone,R., Masson,A.,  
 Mayo,M., Moran,J., Olson,T., Wong,D., Ritland,C.E., Siddiqui,A.,  
 Holt,R., Jones,S., Marra,M., Ellis,B.E., Douglas,C., Ritland,K. and  
 Bohlmann,J.

TITLE  
 The poplar transcriptome: Analysis of expressed sequence tags from  
 multiple cDNA libraries

JOURNAL  
 COMMENT  
 Contact: Joerg Bohlmann  
 Genome BC forest genomics program  
 University of British Columbia  
 Michael Smith Laboratories, 6174 University Boulevard, Rm. 237,  
 Vancouver, British Columbia, Canada, V6T 1Z3  
 Tel: 1-604-822-0282  
 Fax: 1-604-822-2114  
 Email: bohlmann@msl.ubc.ca  
 Plate: WS02523 row: I column: 07  
 High quality sequence stop: 779  
 POLYA=yes.

FEATURES  
 Location/Qualifiers  
 1..779  
 /organism="Populus trichocarpa"  
 /mol\_type="mRNA"  
 /cultivar="Wild clone"  
 /db\_xref="taxon:3694"  
 /clone="WS02523\_107"  
 /sex="Male"  
 /lab\_host="E. coli DH10B T1 phage resistant cells"  
 /clone\_lib="PT-MB-N-A-15"  
 /note="Vector: pBluescript II SK (+) XR; Site 1: EcoRI (5'  
 end of cDNA); Site 2: XhoI (3' end of cDNA); Terminal  
 vegetative buds from 20 year old trees harvested near  
 Corvallis, Oregon on September 19th, 2001. cDNA was  
 prepared from 5 micrograms of mRNA and directionally  
 ligated into the pBluescript II SK (+) XR vector using the  
 pBluescript II XR cDNA Library Construction Kit according  
 to manufacturer's instructions with modifications  
 (Stratagene). Plasmid DNA was then transformed by  
 electroporation into DH10B cells (Invitrogen) for  
 propagation. Normalization was applied according to  
 published methods (Bonaldo M.F. et al. (1996) Genome  
 Research 6(9):791) in order to reduce the abundance of  
 highly expressed transcripts."

Query Match 91.6%; Score 17.4; DB 10; Length 779;  
 Best Local Similarity 94.7%; Pred. No. 2e+03;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCTTTG 19  
 ||||| ||||| |||||  
 Db 660 CTTTACTTCATAGTCTTTG 678



```

RESULT 62
DT483764/c
LOCUS
DEFINITION
779 bp mRNA linear EST 29-AUG-2005
PT-MB-N-A-15 Populus trichocarpa cDNA clone
WS02523.B21_I07_3', mRNA sequence.

ACCESSION
DT483764
VERSION
DT483764.1
KEYWORDS
GI:73881026
SOURCE
Populus trichocarpa (Populus balsamifera subsp. trichocarpa)
ORGANISM
Populus trichocarpa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids I; Malpighiales; Salicaceae; Populus.
1 (bases 1 to 779)
REFERENCE
Ralph,S., Cooper,D., Kolosova,N., Oddy,C., Butterfield,Y.,
Kirkpatrick,R., Liu,J., Palmquist,D., Stott,J., Barber,S., Yang,G.,
Babakaiff,R., Brown-John,M., Chand,S., Featherstone,R., Masson,A.,
Mayo,M., Moran,J., Olson,T., Wong,D., Ritland,C.E., Siddiqui,A.,
Holt,R., Jones,S., Marra,M., Ellis,B.E., Douglas,C., Ritland,K. and
Bohlmann,J.
The poplar transcriptome: Analysis of expressed sequence tags from
multiple cDNA libraries
Unpublished (2004)
JOURNAL
COMMENT
Contact: Joerg Bohlmann
Genome BC forest genomics program
University of British Columbia
Michael Smith Laboratories, 6174 University Boulevard, Rm. 237,
Vancouver, British Columbia, Canada, V6T 1Z3
Tel: 1-604-822-0282
Fax: 1-604-822-2114
Email: bohlmann@mel.ubc.ca
Plate: WS02523 row: I column: 07
High quality sequence stop: 779
POLYA=Yes.

FEATURES
source
1..779
Location/Qualifiers
/organism="Populus trichocarpa"
/mol_type="mRNA"
/cultivar="Wild clone"
/db_xref="taxon:3694"
/clone="WS02523_I07"
/sex="Male"
/lab_host="E. coli DH10B T1 phage resistant cells"
/notes="Vector: pBluescript II SK (+) XR; Site 1: EcoRI (5'
end of cDNA); Site 2: XhoI (3' end of cDNA); Terminal
vegetative buds from 20 year old trees harvested near
Corvallis, Oregon on September 19th, 2001. cDNA was
prepared from 5 micrograms of mRNA and directionally
ligated into the pBluescript II SK (+) XR vector using the
pBluescript II XR cDNA Library Construction Kit according
to manufacturer's instructions with modifications
(Stratagene). Plasmid DNA was then transformed by
electroporation into DH10B cells (Invitrogen) for
propagation. Normalization was applied according to
published methods [Bonaldo M.F. et al. (1996) Genome
Research 6(9):791] in order to reduce the abundance of
highly expressed transcripts."

ORIGIN
Query Match 91.6%; Score 17.4; DB 10; Length 779;
Best Local Similarity 94.7%; Pred. No. 2e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCCTTTG 19
|||||
Db 120 CTTTACTACATAGTCCTTTG 102

RESULT 63
DT483588/c
LOCUS
DEFINITION
789 bp mRNA linear EST 29-AUG-2005
PT-MB-N-A-15 Populus trichocarpa cDNA clone
WS02523.B21_A10_3', mRNA sequence.

ACCESSION
DT483588
VERSION
DT483588.1
KEYWORDS
GI:73880850
SOURCE
Populus trichocarpa (Populus balsamifera subsp. trichocarpa)
ORGANISM
Populus trichocarpa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids I; Malpighiales; Salicaceae; Populus.
1 (bases 1 to 789)
REFERENCE
Ralph,S., Cooper,D., Kolosova,N., Oddy,C., Butterfield,Y.,
Kirkpatrick,R., Liu,J., Palmquist,D., Stott,J., Barber,S., Yang,G.,
Babakaiff,R., Brown-John,M., Chand,S., Featherstone,R., Masson,A.,
Mayo,M., Moran,J., Olson,T., Wong,D., Ritland,C.E., Siddiqui,A.,
Holt,R., Jones,S., Marra,M., Ellis,B.E., Douglas,C., Ritland,K. and
Bohlmann,J.
The poplar transcriptome: Analysis of expressed sequence tags from
multiple cDNA libraries
Unpublished (2004)
JOURNAL
COMMENT
Contact: Joerg Bohlmann
Genome BC forest genomics program
University of British Columbia
Michael Smith Laboratories, 6174 University Boulevard, Rm. 237,
Vancouver, British Columbia, Canada, V6T 1Z3
Tel: 1-604-822-0282
Fax: 1-604-822-2114
Email: bohlmann@mel.ubc.ca
Plate: WS02523 row: A column: 10
High quality sequence stop: 789
POLYA=Yes.

FEATURES
source
1..789
Location/Qualifiers
/organism="Populus trichocarpa"
/mol_type="mRNA"
/cultivar="Wild clone"
/db_xref="taxon:3694"
/clone="WS02523_A10"
/sex="Male"
/lab_host="E. coli DH10B T1 phage resistant cells"
/notes="Vector: pBluescript II SK (+) XR; Site 1: EcoRI (5'
end of cDNA); Site 2: XhoI (3' end of cDNA); Terminal
vegetative buds from 20 year old trees harvested near
Corvallis, Oregon on September 19th, 2001. cDNA was
prepared from 5 micrograms of mRNA and directionally
ligated into the pBluescript II SK (+) XR vector using the
pBluescript II XR cDNA Library Construction Kit according
to manufacturer's instructions with modifications
(Stratagene). Plasmid DNA was then transformed by
electroporation into DH10B cells (Invitrogen) for
propagation. Normalization was applied according to
published methods [Bonaldo M.F. et al. (1996) Genome
Research 6(9):791] in order to reduce the abundance of
highly expressed transcripts."

ORIGIN
Query Match 91.6%; Score 17.4; DB 10; Length 789;
Best Local Similarity 94.7%; Pred. No. 2e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCCTTTG 19
|||||
Db 116 CTTTACTACATAGTCCTTTG 98

RESULT 64
CV258213/c
LOCUS
DEFINITION
802 bp mRNA linear EST 22-SEP-2004
WS0248.B21_O02_PTXD-ICC-N-A-14 Populus trichocarpa x Populus
deltoides cDNA clone WS0248_O02_3', mRNA sequence.
CV258213
ACCESSION
CV258213

```



Fax: 1-604-822-2114  
Email: bohlmann@msl.ubc.ca  
Plate: WS02514 row: E column: 19  
High quality sequence stop: 834  
POLYA=Yes.

# FEATURES

source  
Location/Qualifiers  
1..834  
/organism="Populus trichocarpa"  
/mol\_type="mRNA"  
/cultivar="Wild clone"  
/db\_xref="taxon:3694"  
/clones="WS02514\_E19"  
/sex="Male"  
/lab\_host="E. coli DH10B T1 phage resistant cells"  
/clone\_lib="PT-MB-N-A-15"  
/note="Vector: pBluescript II SK (+) XR; Site 1: EcoRI (5' end of cDNA); Site 2: XhoI (3' end of cDNA); Terminal vegetative buds from 20 year old trees harvested near Corvallis, Oregon on September 19th, 2001. cDNA was prepared from 5 micrograms of mRNA and directionally ligated into the pBluescript II SK (+) XR vector using the pBluescript II XR cDNA Library Construction Kit according to manufacturer's instructions with modifications (Stratagene). Plasmid DNA was then transformed by electroporation into DH10B cells (Invitrogen) for propagation. Normalization was applied according to published methods [Bonaldo M.F. et al. (1996) Genome Research 6(9):791] in order to reduce the abundance of highly expressed transcripts."

# ORIGIN

Query Match 91.6%; Score 17.4; DB 8; Length 834;  
Best Local Similarity 94.7%; Pred. No. 2.1e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 CTTTACTTCATAGTCTTTG 19  
||||| |||||  
Db 120 CTTTACTACATAGTCTTTG 102

RESULT 67  
DT506297/c  
LOCUS DT506297 837 bp mRNA linear EST 29-AUG-2005  
DEFINITION WS0189.C21 D10 PTxD-IL-N-A-9 Populus trichocarpa x Populus deltoides cDNA clone WS0189\_D10 3', mRNA sequence.  
ACCESSION DT506297  
VERSION DT506297.1 GI:73903559  
KEYWORDS EST.  
SOURCE Populus trichocarpa x Populus deltoides  
ORGANISM Populus trichocarpa x Populus deltoides  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids 1; Malpighiales; Salicaceae; Populus.

REFERENCE 1 (bases 1 to 837)  
AUTHORS Ralph,S., Cooper,D., Kolosova,N., Oddy,C., Butterfield,Y., Kirkpatrick,R., Liu,J., Palmquist,D., Scott,J., Barber,S., Yang,G., Babakaiff,R., Brown-John,M., Chand,S., Featherstone,R., Masson,A., Mayo,M., Moran,J., Olson,T., Wong,D., Ritland,C.E., Siddiqui,A., Holt,R., Jones,S., Marra,M., Ellis,B.E., Douglas,C., Ritland,K. and Bohlmann,J.  
The poplar transcriptome: Analysis of expressed sequence tags from multiple cDNA libraries  
Unpublished (2004)  
Contact: Joerg Bohlmann  
Genome BC forest genomics program  
University of British Columbia  
Michael Smith Laboratories, 6174 University Boulevard, Rm. 237,  
Vancouver, British Columbia, Canada, V6T 1Z3  
Tel: 1-604-822-0282  
Fax: 1-604-822-2114  
Email: bohlmann@msl.ubc.ca  
Plate: WS0189 row: D column: 10  
High quality sequence stop: 837

# POLYA=Yes.

FEATURES  
source  
Location/Qualifiers  
1..837  
/organism="Populus trichocarpa x Populus deltoides"  
/mol\_type="mRNA"  
/cultivar="Hil-11"  
/db\_xref="taxon:3695"  
/clones="WS0189\_D10"  
/sex="Male"  
/lab\_host="E. coli DH10B T1 phage resistant cells"  
/clone\_lib="PTxD-IL-N-A-9"  
/note="Vector: pBluescript II SK (+) XR; Site 1: EcoRI (5' end of cDNA); Site 2: XhoI (3' end of cDNA); Sapling trees two metres in height and grown under greenhouse conditions were exposed to continuous feeding by Malacosoma disstria Hubner (forest tent caterpillar) mid-instar larvae caged on the sapling using mesh bags. Mature leaves from within the caged region were collected 2 hours, 12 hours, 24 hours and 48 hours after the onset of treatment. mRNA was isolated from each tissue source independently and equal quantities of mRNA from each tissue were then pooled. cDNA was prepared from 5 micrograms of mRNA and directionally ligated into the pBluescript II SK (+) XR vector using the pBluescript II XR cDNA Library Construction Kit according to manufacturer's instructions with modifications (Stratagene). Plasmid DNA was then transformed by electroporation into DH10B cells (Invitrogen) for propagation. Normalization was applied according to published methods [Bonaldo M.F. et al. (1996) Genome Research 6(9):791] in order to reduce the abundance of highly expressed transcripts."

# ORIGIN

Query Match 91.6%; Score 17.4; DB 10; Length 837;  
Best Local Similarity 94.7%; Pred. No. 2.1e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 CTTTACTTCATAGTCTTTG 19  
||||| |||||  
Db 315 CTTTACTACATAGTCTTTG 297

RESULT 68  
DT487539  
LOCUS DT487539 839 bp mRNA linear EST 29-AUG-2005  
DEFINITION WS02533.B21\_P14 PT-MB-N-A-15 Populus trichocarpa cDNA clone WS02533\_P14 3', mRNA sequence.  
ACCESSION DT487539  
VERSION DT487539.1 GI:73884801  
KEYWORDS EST.  
SOURCE Populus trichocarpa (Populus balsamifera subsp. trichocarpa)  
ORGANISM Populus trichocarpa  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids 1; Malpighiales; Salicaceae; Populus.

REFERENCE 1 (bases 1 to 839)  
AUTHORS Ralph,S., Cooper,D., Kolosova,N., Oddy,C., Butterfield,Y., Kirkpatrick,R., Liu,J., Palmquist,D., Scott,J., Barber,S., Yang,G., Babakaiff,R., Brown-John,M., Chand,S., Featherstone,R., Masson,A., Mayo,M., Moran,J., Olson,T., Wong,D., Ritland,C.E., Siddiqui,A., Holt,R., Jones,S., Marra,M., Ellis,B.E., Douglas,C., Ritland,K. and Bohlmann,J.  
The poplar transcriptome: Analysis of expressed sequence tags from multiple cDNA libraries  
Unpublished (2004)  
Contact: Joerg Bohlmann  
Genome BC forest genomics program  
University of British Columbia  
Michael Smith Laboratories, 6174 University Boulevard, Rm. 237,  
Vancouver, British Columbia, Canada, V6T 1Z3  
Tel: 1-604-822-0282  
Fax: 1-604-822-2114  
Email: bohlmann@msl.ubc.ca

TITLE The poplar transcriptome: Analysis of expressed sequence tags from multiple cDNA libraries  
JOURNAL Unpublished (2004)  
COMMENT Contact: Joerg Bohlmann  
Genome BC forest genomics program  
University of British Columbia  
Michael Smith Laboratories, 6174 University Boulevard, Rm. 237,  
Vancouver, British Columbia, Canada, V6T 1Z3  
Tel: 1-604-822-0282  
Fax: 1-604-822-2114  
Email: bohlmann@msl.ubc.ca

```

Plate: WS02533 row: P column: 14
High quality sequence stop: 839.
FEATURES
    source
        1..839
            /organism="Populus trichocarpa"
            /mol_type="mRNA"
            /cultivar="Wild clone"
            /db_xref="taxon:3694"
            /clone="WS02533_P14"
            /sex="Male"
            /lab_host="E. coli DH10B T1 phage resistant cells"
            /clone_lib="PT-MB-N-A-15"
            /note="Vector: pBluescript II SK (+) XR; Site_1: EcoRI (5'
            end of cDNA); Site_2: XhoI (3' end of cDNA); Terminal
            vegetative buds from 20 year old trees harvested near
            Corvallis, Oregon on September 19th, 2001. cDNA was
            prepared from 5 micrograms of mRNA and directionally
            ligated into the pBluescript II SK (+) XR vector using the
            pBluescript II XR cDNA Library Construction Kit according
            to manufacturer's instructions with modifications
            (Stratagene). Plasmid DNA was then transformed by
            electroporation into DH10B cells (Invitrogen) for
            propagation. Normalization was applied according to
            published methods [Bonaldo M.F. et al. (1996) Genome
            Research 6(9):791] in order to reduce the abundance of
            highly expressed transcripts."
ORIGIN
    Query Match          91.6%; Score 17.4; DB 10; Length 839;
    Best Local Similarity 94.7%; Pred. No. 2.1e+03;
    Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

    QY 1 CTTTACTTCATAGTCCTTTG 19
        ||||| ||||| ||||| |||||
    Db 735 CTTTACTACATAGTCCTTTG 753

RESULT 69
DT487258/c
LOCUS DT487258 844 bp mRNA linear EST 29-AUG-2005
DEFINITION WS02533 B21_D09 PT-MB-N-A-15 Populus trichocarpa cDNA clone
ACCESSION WS02533_D09_3, mRNA sequence.
VERSION DT487258
KEYWORDS EST.
SOURCE Populus trichocarpa (Populus balsamifera subsp. trichocarpa)
ORGANISM Populus trichocarpa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids I; Malpighiales; Salicaceae; Populus.
1 (bases 1 to 844)
Ralph,S., Cooper,D., Kolosova,N., Oddy,C., Butterfield,Y.,
Kirkpatrick,R., Liu,J., Palmquist,D., Stott,J., Barber,S., Yang,G.,
Babakaiff,R., Brown-John,M., Chand,S., Featherstone,R., Maason,A.,
Mayo,M., Moran,J., Olson,T., Wong,D., Ritland,C.E., Siddiqui,A.,
Holt,R., Jones,S., Marra,M., Ellis,B.E., Douglas,C., Ritland,K. and
Bohlmann,J.
The poplar transcriptome: Analysis of expressed sequence tags from
multiple cDNA libraries
Unpublished (2004)
Contact: Joerg Bohlmann
Genome BC forest genomics program
University of British Columbia
Michael Smith Laboratories, 6174 University Boulevard, Rm. 237,
Vancouver, British Columbia, Canada, V6T 1Z3
Tel: 1-604-822-0282
Fax: 1-604-822-2114
Email: bohlmann@msl.ubc.ca
Plate: WS02533 row: D column: 09
High quality sequence stop: 844
POLYA=Yes. Location/Qualifiers
    source
        1..844

Plate: WS02533 row: P column: 14
High quality sequence stop: 839.
FEATURES
    source
        1..839
            /organism="Populus trichocarpa"
            /mol_type="mRNA"
            /cultivar="Wild clone"
            /db_xref="taxon:3694"
            /clone="WS02533_P14"
            /sex="Male"
            /lab_host="E. coli DH10B T1 phage resistant cells"
            /clone_lib="PT-MB-N-A-15"
            /note="Vector: pBluescript II SK (+) XR; Site_1: EcoRI (5'
            end of cDNA); Site_2: XhoI (3' end of cDNA); Terminal
            vegetative buds from 20 year old trees harvested near
            Corvallis, Oregon on September 19th, 2001. cDNA was
            prepared from 5 micrograms of mRNA and directionally
            ligated into the pBluescript II SK (+) XR vector using the
            pBluescript II XR cDNA Library Construction Kit according
            to manufacturer's instructions with modifications
            (Stratagene). Plasmid DNA was then transformed by
            electroporation into DH10B cells (Invitrogen) for
            propagation. Normalization was applied according to
            published methods [Bonaldo M.F. et al. (1996) Genome
            Research 6(9):791] in order to reduce the abundance of
            highly expressed transcripts."
ORIGIN
    Query Match          91.6%; Score 17.4; DB 10; Length 839;
    Best Local Similarity 94.7%; Pred. No. 2.1e+03;
    Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

    QY 1 CTTTACTTCATAGTCCTTTG 19
        ||||| ||||| ||||| |||||
    Db 735 CTTTACTACATAGTCCTTTG 753

RESULT 70
CV254849/c
LOCUS CV254849 852 bp mRNA linear EST 22-SEP-2004
DEFINITION WS02411 B21_F21 PTxD-ICC-N-A-14 Populus trichocarpa x Populus
deltoides cDNA clone WS02411_F21_3, mRNA sequence.
ACCESSION CV254849
VERSION CV254849.1 GI:52507824
KEYWORDS EST.
SOURCE Populus trichocarpa x Populus deltoides
ORGANISM Populus trichocarpa x Populus deltoides
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids I; Malpighiales; Salicaceae; Populus.
1 (bases 1 to 852)
Ralph,S., Cooper,D., Kolosova,N., Oddy,C., Butterfield,Y.,
Kirkpatrick,R., Liu,J., Palmquist,D., Stott,J., Barber,S., Yang,G.,
Babakaiff,R., Brown-John,M., Chand,S., Featherstone,R., Maason,A.,
Mayo,M., Moran,J., Olson,T., Wong,D., Ritland,C.E., Siddiqui,A.,
Holt,R., Jones,S., Marra,M., Ellis,B.E., Douglas,C., Ritland,K. and
Bohlmann,J.
The poplar transcriptome: Analysis of expressed sequence tags from
multiple cDNA libraries
Unpublished (2004)
Contact: Joerg Bohlmann
Genome BC forest genomics program
University of British Columbia
Michael Smith Laboratories, 6174 University Boulevard, Rm. 237,
Vancouver, British Columbia, Canada, V6T 1Z3
Tel: 1-604-822-0282
Fax: 1-604-822-2114
Email: bohlmann@msl.ubc.ca
Plate: WS02411 row: F column: 21
High quality sequence stop: 852
POLYA=Yes. Location/Qualifiers
    source
        1..852
            /organism="Populus trichocarpa x Populus deltoides"
            /mol_type="mRNA"
            /cultivar="H11-11"
            /db_xref="taxon:3695"

```

```

/clonew="WS02411_F21"
/sex="Male"
/lab host="E. coli DH10B T1 phage resistant cells"
/clonelib="PTXD-ICC-N-A-14"
/notes="vector: pBluescript II SK (+) XR; Site 1: EcoRI (5' end of cDNA); Site 2: XhoI (3' end of cDNA); Cultured cells (de Sa MM et al. (1992) Plant Physiology 98:728-737) were grown in media (45mL) supplemented with either 50uM salicylic acid, 50uM benzothiadiazole, 50uM methyl jasmonate, 20uM chitosan or 200uL of Pollacia radiosa extract. Cells were harvested after a 3 hour treatment, along with untreated control cells. mRNA was isolated from each tissue source independently and equal quantities of mRNA from each tissue were then pooled. cDNA was prepared from 5 micrograms of mRNA and directionally ligated into the pBluescript II SK (+) XR vector using the pBluescript II XR cDNA Library Construction Kit according to manufacturer's instructions with modifications (Stratagene). Plasmid DNA was then transformed by electroporation into DH10B cells (Invitrogen) for propagation. Normalization was applied according to published methods [Bonaldo M.F. et al. (1996) Genome Research 6(9):791] in order to reduce the abundance of highly expressed transcripts."

```

## ORIGIN

```

Query Match      91.6%; Score 17.4; DB 8; Length 852;
Best Local Similarity 94.7%; Pred. No. 2.1e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 CTTTACTTCATAGTCCTTTG 19
||||| ||||| ||||| |||||
DB 120 CTTTACTACATAGTCCTTTG 102

```

## RESULT 71

```

DT489385/c
LOCUS      859 bp mRNA linear EST 29-AUG-2005
DEFINITION WS02542 B21 B08 PT-MB-N-A-15 Populus trichocarpa cDNA clone
VERSION    WS02542 B08 3', mRNA sequence.
KEYWORDS   DT489385.1 GI:73886647
SOURCE     EST.
ORGANISM   Populus trichocarpa (Populus balsamifera subsp. trichocarpa)

```

## REFERENCE

```

AUTHORS    Ralph,S., Cooper,D., Kolosova,N., Oddy,C., Butterfield,Y., Kirkpatrick,R., Liu,J., Palmquist,D., Stott,J., Barber,S., Yang,G., Babakaiff,R., Brown-John,M., Chand,S., Featherstone,R., Masson,A., Mayo,M., Moran,J., Olson,T., Wong,D., Ritland,C.E., Siddiqui,A., Holt,R., Jones,S., Marra,M., Ellis,B.E., Douglas,C., Ritland,K. and Bohlmann,J.
TITLE      The poplar transcriptome: Analysis of expressed sequence tags from multiple cDNA libraries
JOURNAL    Unpublished (2004)
COMMENT    Contact: Joerg Bohlmann
Genome BC forest genomics program
University of British Columbia
Michael Smith Laboratories, 6174 University Boulevard, Rm. 237, Vancouver, British Columbia, Canada, V6T 1Z3
Tel: 1-604-822-0282
Fax: 1-604-822-2114
Email: bohlmann@msl.ubc.ca
Plate: WS02542 row: B column: 08
High quality sequence stop: 859
POLYA=Yes.

```

## FEATURES

```

source
1..859
/organism="Populus trichocarpa"
/mol_type="mRNA"

```

```

/cultivar="Wild clone"
/db xref="taxon:3694"
/clonew="WS02542_B08"
/sex="Male"
/lab host="E. coli DH10B T1 phage resistant cells"
/clonelib="PT-MB-N-A-15"
/notes="vector: pBluescript II SK (+) XR; Site 1: EcoRI (5' end of cDNA); Site 2: XhoI (3' end of cDNA); Terminal vegetative buds from 20 year old trees harvested near Corvallis, Oregon on September 19th, 2001. cDNA was prepared from 5 micrograms of mRNA and directionally ligated into the pBluescript II SK (+) XR vector using the pBluescript II XR cDNA Library Construction Kit according to manufacturer's instructions with modifications (Stratagene). Plasmid DNA was then transformed by electroporation into DH10B cells (Invitrogen) for propagation. Normalization was applied according to published methods [Bonaldo M.F. et al. (1996) Genome Research 6(9):791] in order to reduce the abundance of highly expressed transcripts."

```

## ORIGIN

```

Query Match      91.6%; Score 17.4; DB 10; Length 859;
Best Local Similarity 94.7%; Pred. No. 2.1e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 CTTTACTTCATAGTCCTTTG 19
||||| ||||| ||||| |||||
DB 121 CTTTACTACATAGTCCTTTG 103

```

## RESULT 72

```

CV245752/c
LOCUS      871 bp mRNA linear EST 22-SEP-2004
DEFINITION WS0259 B21 E13 PT-MB-N-A-15 Populus trichocarpa cDNA clone
VERSION    WS0259 E13 3', mRNA sequence.
KEYWORDS   CV245752
SOURCE     EST.
ORGANISM   Populus trichocarpa (Populus balsamifera subsp. trichocarpa)

```

## REFERENCE

```

AUTHORS    Ralph,S., Cooper,D., Kolosova,N., Oddy,C., Butterfield,Y., Kirkpatrick,R., Liu,J., Palmquist,D., Stott,J., Barber,S., Yang,G., Babakaiff,R., Brown-John,M., Chand,S., Featherstone,R., Masson,A., Mayo,M., Moran,J., Olson,T., Wong,D., Ritland,C.E., Siddiqui,A., Holt,R., Jones,S., Marra,M., Ellis,B.E., Douglas,C., Ritland,K. and Bohlmann,J.
TITLE      The poplar transcriptome: Analysis of expressed sequence tags from multiple cDNA libraries
JOURNAL    Unpublished (2004)
COMMENT    Contact: Joerg Bohlmann
Genome BC forest genomics program
University of British Columbia
Michael Smith Laboratories, 6174 University Boulevard, Rm. 237, Vancouver, British Columbia, Canada, V6T 1Z3
Tel: 1-604-822-0282
Fax: 1-604-822-2114
Email: bohlmann@msl.ubc.ca
Plate: WS0259 row: E column: 13
High quality sequence stop: 871
POLYA=Yes.

```

## FEATURES

```

source
1..871
/organism="Populus trichocarpa"
/mol_type="mRNA"
/cultivar="Wild clone"
/db xref="taxon:3694"
/clonew="WS0259_E13"
/sex="Male"

```

/lab\_host="E. coli DH10B T1 phage resistant cells"  
 /clone\_lib="PT-MB-N-A-15"  
 /notes="Vector: pBluescript II SK (+) XR; Site 1: EcoRI (5' end of cDNA); Site 2: XhoI (3' end of cDNA); Terminal vegetative buds from 20 year old trees harvested near Corvallis, Oregon on September 19th, 2001. cDNA was prepared from 5 micrograms of mRNA and directionally ligated into the pBluescript II SK (+) XR vector using the pBluescript II XR cDNA Library Construction Kit according to manufacturer's instructions with modifications (Stratagene). Plasmid DNA was then transformed by electroporation into DH10B cells (Invitrogen) for propagation. Normalization was applied according to published methods [Bonaldo M.F. et al. (1996) Genome Research 6(9):791] in order to reduce the abundance of highly expressed transcripts."

## ORIGIN

Query Match 91.6%; Score 17.4; DB 8; Length 871;  
 Best Local Similarity 94.7%; Pred. No. 2.1e+03;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCCTTTG 19  
 ||||| ||||| ||||| |||||  
 Db 117 CTTTACTACATAGTCCTTTG 99

## RESULT 73

DT522272/c 877 bp mRNA linear EST 30-AUG-2005  
 LOCUS WS02036.B21\_I14 PTXN-IB-N-A-11 Populus trichocarpa x Populus nigra  
 DEFINITION cDNA clone WS02036\_I14 3', mRNA sequence.

ACCESSION DT522272  
 VERSION DT522272.1 GI:73938962

KEYWORDS EST  
 SOURCE Populus trichocarpa x Populus nigra

ORGANISM Populus trichocarpa x Populus nigra  
 Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
 rosids; eustosids I; Malpighiales; Salicaceae; Populus.

REFERENCE 1 (bases 1 to 877)

AUTHORS Ralph,S., Cooper,D., Kolosova,N., Oddy,C., Butterfield,Y., Kirkpatrick,R., Liu,J., Palmquist,D., Stott,J., Barber,S., Yang,G., Babakaiff,R., Brown-John,M., Chand,S., Featherstone,R., Masson,A., Mayo,M., Moran,J., Olson,T., Wong,D., Ritland,C.E., Siddiqui,A., Holt,R., Jones,S., Marra,M., Ellis,B.E., Douglas,C., Ritland,K. and Bohlmann,J.

TITLE The poplar transcriptome: Analysis of expressed sequence tags from multiple cDNA libraries

JOURNAL Unpublished (2004)  
 COMMENT Contact: Joerg Bohlmann  
 Genome BC forest genomics program  
 University of British Columbia  
 Michael Smith Laboratories, 6174 University Boulevard, Rm. 237,  
 Vancouver, British Columbia, Canada, V6T 1Z3

Tel: 1-604-822-0282

Fax: 1-604-822-2114

Email: bohlmann@msl.ubc.ca

Plate: WS02036 row: I column: 14

High quality sequence stop: 877

POLYA=Yes.

FEATURES  
source

Location/Qualifiers  
 1..877  
 /organism="Populus trichocarpa x Populus nigra"  
 /mol\_type="mRNA"  
 /cultivar="NXM6"  
 /db\_xref="taxon:293756"  
 /clone="WS02036\_I14"  
 /sex="Male"  
 /lab\_host="E. coli DH10B T1 phage resistant cells"  
 /clone\_lib="PTXN-IB-N-A-11"  
 /notes="Vector: pBluescript II SK (+) XR; Site 1: EcoRI (5' end of cDNA); Site 2: XhoI (3' end of cDNA); Sapling trees

two metres in height and grown under greenhouse conditions were exposed to continuous feeding by Cryptorhynchus lapathi (poplar and willow borer) adults caged on the sapling using mesh bags. Bark (with phloem and cambium attached) from within the caged region was harvested 2 hours, 6 hours and 48 hours after the onset of treatment. mRNA was isolated from each tissue source independently and equal quantities of mRNA from each tissue were then pooled. cDNA was prepared from 5 micrograms of mRNA and directionally ligated into the pBluescript II SK (+) XR vector using the pBluescript II XR cDNA Library Construction Kit according to manufacturer's instructions with modifications (Stratagene). Plasmid DNA was then transformed by electroporation into DH10B cells (Invitrogen) for propagation. Normalization was applied according to published methods [Bonaldo M.F. et al. (1996) Genome Research 6(9):791] in order to reduce the abundance of highly expressed transcripts."

## ORIGIN

Query Match 91.6%; Score 17.4; DB 10; Length 877;  
 Best Local Similarity 94.7%; Pred. No. 2.1e+03;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCCTTTG 19  
 ||||| ||||| ||||| |||||  
 Db 116 CTTTACTACATAGTCCTTTG 98

## RESULT 74

DT521741/c 887 bp mRNA linear EST 30-AUG-2005  
 LOCUS WS02035.B21\_B16 PTXN-IB-N-A-11 Populus trichocarpa x Populus nigra  
 DEFINITION cDNA clone WS02035\_B16 3', mRNA sequence.

ACCESSION DT521741  
 VERSION DT521741.1 GI:73938431

KEYWORDS EST  
 SOURCE Populus trichocarpa x Populus nigra

ORGANISM Populus trichocarpa x Populus nigra  
 Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
 rosids; eustosids I; Malpighiales; Salicaceae; Populus.

REFERENCE 1 (bases 1 to 887)

AUTHORS Ralph,S., Cooper,D., Kolosova,N., Oddy,C., Butterfield,Y., Kirkpatrick,R., Liu,J., Palmquist,D., Stott,J., Barber,S., Yang,G., Babakaiff,R., Brown-John,M., Chand,S., Featherstone,R., Masson,A., Mayo,M., Moran,J., Olson,T., Wong,D., Ritland,C.E., Siddiqui,A., Holt,R., Jones,S., Marra,M., Ellis,B.E., Douglas,C., Ritland,K. and Bohlmann,J.

TITLE The poplar transcriptome: Analysis of expressed sequence tags from multiple cDNA libraries

JOURNAL Unpublished (2004)  
 COMMENT Contact: Joerg Bohlmann  
 Genome BC forest genomics program  
 University of British Columbia

Michael Smith Laboratories, 6174 University Boulevard, Rm. 237,  
 Vancouver, British Columbia, Canada, V6T 1Z3

Tel: 1-604-822-0282

Fax: 1-604-822-2114

Email: bohlmann@msl.ubc.ca

Plate: WS02035 row: B column: 16

High quality sequence stop: 887

POLYA=Yes.

FEATURES  
source

Location/Qualifiers  
 1..887  
 /organism="Populus trichocarpa x Populus nigra"  
 /mol\_type="mRNA"  
 /cultivar="NXM6"  
 /db\_xref="taxon:293756"  
 /clone="WS02035\_B16"  
 /sex="Male"  
 /lab\_host="E. coli DH10B T1 phage resistant cells"  
 /clone\_lib="PTXN-IB-N-A-11"



**AUTHORS** Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.

**TITLE** Maize Genomics Consortium  
**JOURNAL** Unpublished (2003)  
**COMMENT** Other\_GSSs: PUIDG50TB  
Contact: Cathy Whitelaw

TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-838-5843

Fax: 301-838-0208  
Email: whitelaw@tigr.org

Seq primer: TF  
Class: sheared ends.

**FEATURES**  
source  
1. .955  
Location/Qualifiers  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone\_lib="ZMMBTA0557103"  
/clone\_lib="ZM 0.6 1.0 KB"  
/note="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high  
Col selected genomic DNA library"

**ORIGIN**

Query Match 91.6%; Score 17.4; DB 12; Length 955;  
Best Local Similarity 94.7%; Pred. No. 2.1e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 CTTTACTTCATAGTCTTTG 19  
|||||  
Db 917 CTTTACTTCAAGTCTTTG 899

**RESULT 78**

**CG330802**  
**LOCUS** CG2BM89TV ZM 0.7 1.5 KB Zea mays genomic clone ZMMBMA0753P09,  
**DEFINITION** genomic survey sequence.

**ACCESSION** CG330802  
**VERSION** CG330802.1 GI:34248068  
**KEYWORDS** GSS.

**SOURCE**  
Zea mays  
**ORGANISM**

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 976)

Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,  
Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,  
Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.

Consortium for Maize Genomics

Unpublished (2002)

**COMMENT** Other\_GSSs: OG2BM89TH

Contact: Cathy Whitelaw

TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-838-5843

Fax: 301-838-0208  
Email: whitelaw@tigr.org

Seq primer: TF  
Class: methylation filtered.

**FEATURES**  
source  
1. .976  
Location/Qualifiers  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone\_lib="ZMMBMA0753P09"  
/clone\_lib="ZM 0.7 1.5 KB"  
/note="Vector: pBCSk-; Site 1: HincII; 0.7-1.5 kb  
methylation filtered genomic DNA library"

**ORIGIN**

Query Match 91.6%; Score 17.4; DB 12; Length 976;  
Best Local Similarity 94.7%; Pred. No. 2.1e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 CTTTACTTCATAGTCTTTG 19  
|||||  
Db 84 CTTTACTTCAAGTCTTTG 102

**RESULT 79**

**CB828264/c**  
**LOCUS** CB828264  
**DEFINITION** japonicus Lotus japonicus nodule library 5 and 7 week-old Lotus  
japonicus cDNA 5', mRNA sequence.

**ACCESSION** CB828264  
**VERSION** CB828264.1 GI:29968030  
**KEYWORDS** EST.

**SOURCE**  
Lotus japonicus  
**ORGANISM**

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
rosids; eurosids 1; Fabales; Fabaceae; Papilionoideae; Loteae;  
Lotus.

1 (bases 1 to 459)

Colebatch,G., Freund,S., Trevasakis,B and Udvardi,M.

Lotus japonicus root nodule ESTs: tools for functional genomics

Unpublished (2000)

**COMMENT** Contact: Udvardi MK

Molecular Plant Nutrition

Max Planck Institute of Molecular Plant Physiology

Am Muehlenberg 1, 14476 Golm, Germany

Fax: 49 331 567 8250

Email: udvardi@pimp-golm.mpg.de

Seq primer: T7

High quality sequence stop: 459.

**FEATURES**

source  
1. .459  
Location/Qualifiers  
/organism="Lotus japonicus"  
/mol\_type="mRNA"  
/cultivar="Gifu (B-129)"  
/db\_xref="taxon:34305"  
/dev\_stage="5 and 7 week-old plants"  
/clone\_lib="Lotus japonicus nodule library 5 and 7  
week-old"  
/note="Organ: Nodule; Vector: pSPORT1; Site 1: SalI;  
Site 2: NotI; The library was prepared using mRNA  
extracted from nodules of 5 and 7 week-old Lotus plants.  
Nodules were induced by, and contained Mesorhizobium  
strain R7A."

**ORIGIN**

Query Match 89.5%; Score 17; DB 4; Length 459;  
Best Local Similarity 100.0%; Pred. No. 3e+03;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 TTACTTCATAGTCTTTG 19  
|||||  
Db 372 TTACTTCATAGTCTTTG 356

**RESULT 80**

**AQ431031/c**  
**LOCUS** AQ431031  
**DEFINITION** HS 5087 A2 All T7A RPCI-11 Human Male BAC Library Homo sapiens  
genomic clone Plate=663 Col=22 Row=A, genomic survey sequence.

**ACCESSION** AQ431031  
**VERSION** AQ431031.1 GI:4541366  
**KEYWORDS** GSS.

**SOURCE**  
Homo sapiens (human)  
**ORGANISM**

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;



REFERENCE  
AUTHORS  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
1 (bases 1 to 470)  
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
Hood,L.

TITLE  
Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome

JOURNAL  
PUBMED  
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)

COMMENT  
10449764  
Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center  
University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Clones are derived from the human BAC library RPCI-11. For BAC

library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (http://bacpac.med.buffalo.edu/ordering.bac.htm)

or from Research Genetics (info@resgen.com). BAC end Web Server:

http://www.htesc.washington.edu

Plate: 663 row: A column: 22

Seq primer: T7

Class: BAC ends

High quality sequence stop: 470.

Location/Qualifiers

1. .470

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/clone="Plate:663 Col=22 Row=A"

/sex="male"

/clone\_lib="RPCI-11 Human Male BAC Library"

/note="Vector: pBACE3.6; Site 1: EcoRI; Site 2: EcoRI;

Male blood DNA was isolated from one randomly chosen donor

and partially digested with a combination of EcoRI and

EcoRI Methylase. Size selected DNA was cloned into the

pBACE3.6 vector at EcoRI sites"

ORIGIN

Query Match 89.5%; Score 17; DB 11; Length 470;

Best Local Similarity 100.0%; Pred. No. 3e+03;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 TTACTTCATAGTCTTTG 19

|||||

Db 376 TTACTTCATAGTCTTTG 360

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

Fax: 49 331 567 8250

Email: udvardi@pimp-golm.mpg.de

Seq primer: T7

High quality sequence stop: 552.

Location/Qualifiers

1. .552

/organism="Lotus japonicus"

/mol\_type="mRNA"

/cultivar="Gifu (B-129)"

/db\_xref="taxon:34305"

/dev\_stage="5 and 7 week-old plants"

/clone\_lib="Lotus japonicus nodule library 5 and 7

week-old"

/note="Organ: Nodule; Vector: pSPORT1; Site 1: SalI;

Site 2: NotI; The library was prepared using mRNA

extracted from nodules of 5 and 7 week-old Lotus plants.

Nodules were induced by, and contained Mesorhizobium

strain R7A."

ORIGIN

Query Match 89.5%; Score 17; DB 4; Length 552;

Best Local Similarity 100.0%; Pred. No. 3.1e+03;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 TTACTTCATAGTCTTTG 19

|||||

Db 44 TTACTTCATAGTCTTTG 28

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

FEATURES

source

1. .552

/organism="Lotus japonicus"

/mol\_type="mRNA"

/cultivar="Gifu (B-129)"

/db\_xref="taxon:34305"

/dev\_stage="5 and 7 week-old plants"

/clone\_lib="Lotus japonicus nodule library 5 and 7

week-old"

/note="Organ: Nodule; Vector: pSPORT1; Site 1: SalI;

Site 2: NotI; The library was prepared using mRNA

extracted from nodules of 5 and 7 week-old Lotus plants.

Nodules were induced by, and contained Mesorhizobium

strain R7A."

ORIGIN

Query Match 89.5%; Score 17; DB 4; Length 552;

Best Local Similarity 100.0%; Pred. No. 3.1e+03;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 TTACTTCATAGTCTTTG 19

|||||

Db 44 TTACTTCATAGTCTTTG 28

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

FEATURES

source

1. .578

/organism="Mus musculus"

/mol\_type="genomic DNA"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUGC1M0308C10"

/sex="Male"

/lab\_host="E. Coli strain XL10-Gold, Ti-resistant, F-"

/clone\_lib="Mouse 10kb plasmid UUGC1M library"

ORIGIN

Query Match 89.5%; Score 17; DB 4; Length 552;

Best Local Similarity 100.0%; Pred. No. 3.1e+03;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 TTACTTCATAGTCTTTG 19

|||||

Db 44 TTACTTCATAGTCTTTG 28

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (G114732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

ORIGIN  
Query Match 89.5%; Score 17; DB 11; Length 578;  
Best Local Similarity 100.0%; Pred. No. 3.1e+03;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TTTACTTCATAGCTTT 18  
|||||  
DB 429 TTTACTTCATAGCTTT 445

RESULT 83  
CT259201  
LOCUS  
DEFINITION  
Sus scrofa genomic clone CH242-238M17, genomic survey sequence.  
ACCESSION  
CT259201  
VERSION  
CT259201.1 GI:79715797  
KEYWORDS  
GSS.  
SOURCE  
Sus scrofa (pig)  
ORGANISM  
Sus scrofa  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae;  
Sus.

REFERENCE  
1 (bases 1 to 811)  
Humphray, S.J., Plumb, R.W. and Durham, J.L.  
Direct Submission  
Submitted (01-NOV-2005) The Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk) Unpublished  
This sequence was generated from the SP6 end of BAC 238M17. 238M17 is part of the CHOKI-242 BAC Library created by P. de Jong. Further details: [http://www.sanger.ac.uk/Projects/S\\_scrofa/](http://www.sanger.ac.uk/Projects/S_scrofa/).

FEATURES  
source  
1 . 811  
/organism="Sus scrofa"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9823"  
/clone="CH242-238M17"  
/tissue\_type="White blood cells"  
/notes="vector pTARBAC1.3\_BamHI sex female"

ORIGIN  
Query Match 89.5%; Score 17; DB 14; Length 811;  
Best Local Similarity 100.0%; Pred. No. 3.2e+03;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 TTACTTCATAGCTTTG 19  
|||||  
DB 450 TTACTTCATAGCTTTG 466

RESULT 84  
BI819265/c

LOCUS  
DEFINITION  
BI819265  
603034694F1 NIH\_MGC\_115 Homo sapiens cDNA clone IMAGE:5175718 5', mRNA sequence.

ACCESSION  
BI819265  
VERSION  
BI819265.1 GI:15930815  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens

REFERENCE  
1 (bases 1 to 820)  
NIH-MGC <http://mgi.nci.nih.gov/>.  
NATIONAL INSTITUTES OF HEALTH, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs@mail.nih.gov](mailto:cgabbs@mail.nih.gov)  
Tissue Procurement: Life Technologies, Inc.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Plate: LLAM11437 row: 1 column: 23  
High quality sequence stop: 712.

FEATURES  
source  
1 . 820  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5175718"  
/lab\_host="DH10B"  
/clone\_lib="NIH MGC 115"  
/notes="Organ: pooled brain, lung, testis; Vector: PCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA source anonymous pool of 6 male brains, age range 23-27; 1 male lung, age 27; and 1 male testis, age 69. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.8 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 021. Note: this is a NIH\_MGC Library."

ORIGIN  
Query Match 89.5%; Score 17; DB 2; Length 820;  
Best Local Similarity 100.0%; Pred. No. 3.2e+03;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGCTTT 17  
|||||  
DB 798 CTTTACTTCATAGCTTT 782

RESULT 85  
DU846802  
LOCUS  
DEFINITION  
DU846802  
34821 Tomato HindIII BAC Library Lycopersicon esculentum genomic clone LE HBa0106N11 5, genomic survey sequence.

ACCESSION  
DU846802  
VERSION  
DU846802.1 GI:84054989  
KEYWORDS  
GSS.  
SOURCE  
Lycopersicon esculentum (Solanum lycopersicum)  
ORGANISM  
Lycopersicon esculentum  
Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.

REFERENCE  
1 (bases 1 to 884)  
Mueller, L.A., Bueis, R.M., Wang, Y., Tanksley, S.D., Giovannoni, J.J., Van Eck, J. and Stack, S.  
BAC end sequencing from three Solanum lycopersicon libraries  
Unpublished (2005)

ORIGIN  
Query Match 89.5%; Score 17; DB 2; Length 820;  
Best Local Similarity 100.0%; Pred. No. 3.2e+03;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGCTTT 17  
|||||  
DB 798 CTTTACTTCATAGCTTT 782

RESULT 85  
DU846802  
LOCUS  
DEFINITION  
DU846802  
34821 Tomato HindIII BAC Library Lycopersicon esculentum genomic clone LE HBa0106N11 5, genomic survey sequence.

ACCESSION  
DU846802  
VERSION  
DU846802.1 GI:84054989  
KEYWORDS  
GSS.  
SOURCE  
Lycopersicon esculentum (Solanum lycopersicum)  
ORGANISM  
Lycopersicon esculentum  
Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.

REFERENCE  
1 (bases 1 to 884)  
Mueller, L.A., Bueis, R.M., Wang, Y., Tanksley, S.D., Giovannoni, J.J., Van Eck, J. and Stack, S.  
BAC end sequencing from three Solanum lycopersicon libraries  
Unpublished (2005)

ORIGIN  
Query Match 89.5%; Score 17; DB 2; Length 820;  
Best Local Similarity 100.0%; Pred. No. 3.2e+03;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGCTTT 17  
|||||  
DB 798 CTTTACTTCATAGCTTT 782

RESULT 85  
DU846802  
LOCUS  
DEFINITION  
DU846802  
34821 Tomato HindIII BAC Library Lycopersicon esculentum genomic clone LE HBa0106N11 5, genomic survey sequence.

ACCESSION  
DU846802  
VERSION  
DU846802.1 GI:84054989  
KEYWORDS  
GSS.  
SOURCE  
Lycopersicon esculentum (Solanum lycopersicum)  
ORGANISM  
Lycopersicon esculentum  
Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.

REFERENCE  
1 (bases 1 to 884)  
Mueller, L.A., Bueis, R.M., Wang, Y., Tanksley, S.D., Giovannoni, J.J., Van Eck, J. and Stack, S.  
BAC end sequencing from three Solanum lycopersicon libraries  
Unpublished (2005)

ORIGIN  
Query Match 89.5%; Score 17; DB 2; Length 820;  
Best Local Similarity 100.0%; Pred. No. 3.2e+03;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGCTTT 17  
|||||  
DB 798 CTTTACTTCATAGCTTT 782

## COMMENT

Other GSSs: 34820  
 Contact: Lukas Mueller  
 Tanksley Lab, Dept. of Plant Breeding  
 Cornell University  
 251 Emerson Hall, Ithaca, NY 14853, USA  
 Tel: 607-255-6557  
 Fax: 607-255-6683  
 Email: sgn-feedback@sgn.cornell.edu  
 Plate: 106 row: N column: 11  
 Seq primer: T7  
 Class: BAC ends  
 High quality sequence start: 29  
 High quality sequence stop: 716.

## FEATURES

source  
 1..884  
 /organism="Lycopersicon esculentum"  
 /mol\_type="genomic DNA"  
 /cultivar="Heinz 1706"  
 /db\_xref="taxon:4081"  
 /clone="LE\_HBa0106N11"  
 /lab\_host="E. coli"  
 /clone\_lib="Tomato HindIII BAC Library"  
 /note="vector: pBelOBAC11; Site\_1: HindIII"

## ORIGIN

Query Match 89.5%; Score 17; DB 14; Length 884;  
 Best Local Similarity 100.0%; Pred. No. 3.2e+03;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 TTACTTCATAGTCTTTG 19

|||||

Db 664 TTACTTCATAGTCTTTG 680

## RESULT 86

DU870629

LOCUS

DEFINITION 99008 Tomato HindIII BAC Library Lycopersicon esculentum genomic clone LE\_HBa0132G17 3, genomic survey sequence.

ACCESSION DU870629

VERSION DU870629.1

KEYWORDS GI:84079172

SOURCE GSS.

ORGANISM Lycopersicon esculentum (Solanum lycopersicum)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.

REFERENCE 1 (bases 1 to 894)

AUTHORS Mueller, L.A., Buel, R.M., Wang, Y., Tanksley, S.D., Giovannoni, J.J.,

Van Eck, J. and Stack, S.

TITLE BAC end sequencing from three Solanum lycopersicon libraries

JOURNAL Unpublished (2005)

COMMENT Other GSSs: 99009

Contact: Lukas Mueller  
 Tanksley Lab, Dept. of Plant Breeding  
 Cornell University  
 251 Emerson Hall, Ithaca, NY 14853, USA  
 Tel: 607-255-6557  
 Fax: 607-255-6683  
 Email: sgn-feedback@sgn.cornell.edu  
 Insert Length: 110966 Std Error: 0.00  
 Plate: 132 row: G column: 17  
 Seq primer: SP6  
 Class: BAC ends  
 High quality sequence start: 21  
 High quality sequence stop: 396.

## FEATURES

source  
 1..894  
 /organism="Lycopersicon esculentum"  
 /mol\_type="genomic DNA"  
 /cultivar="Heinz 1706"  
 /db\_xref="taxon:4081"  
 /clone="LE\_HBa0132G17"

## ORIGIN

Query Match 89.5%; Score 17; DB 14; Length 894;  
 Best Local Similarity 100.0%; Pred. No. 3.2e+03;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 TTACTTCATAGTCTTTG 19

|||||

Db 24 TTACTTCATAGTCTTTG 40

## RESULT 87

BU931252/c

LOCUS

DEFINITION 907 bp mRNA linear EST 18-OCT-2002 AGENCOURT 10478234 NIH MGC\_109 Homo sapiens cDNA clone IMAGE:6671391 5', mRNA sequence.

ACCESSION BU931252

VERSION BU931252.1

KEYWORDS GI:24120071

SOURCE EST.

ORGANISM Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 907)

AUTHORS NIH-MGC http://mgi.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: sgapbs-x@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLCM2950 row: d column: 15

High quality sequence stop: 550.

FEATURES Location/Qualifiers

source 1..907

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:6671391"

/tissue\_type="teratocarcinoma, cell line"

/lab\_host="DH10B (phage-resistant)"

/clone\_lib="NIH\_MGC\_109"

/note="Organ: ovary; Vector: pOTB7; Site\_1: EcoRI; Site\_2:

XhoI; cDNA made by oligo-dT priming. Directionally cloned

into EcoRI/XhoI sites using the following 5' adaptor:

GGCAGCAG(G). Library constructed by Ling Hong in the

laboratory of Gerald M. Rubin (University of California,

Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and

Superscript II RT (Life Technologies). Note: this is a

NIH\_MGC Library."

## ORIGIN

Query Match 89.5%; Score 17; DB 3; Length 907;  
 Best Local Similarity 100.0%; Pred. No. 3.2e+03;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TTACTTCATAGTCTTT 18

|||||

Db 775 TTACTTCATAGTCTTT 759

## RESULT 88

AK171788/c

LOCUS

DEFINITION 2921 bp mRNA linear HTC 21-SEP-2005

DEFINITION	Mus musculus activated spleen cDNA, RIKEN full-length enriched library, clone:F83001E06 product:transformer-2 protein homolog [Mus musculus], full insert sequence.
ACCESSION	AK171788
VERSION	AK171788.1 GI:74152821
KEYWORDS	HTC; CAP trapper.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	1 Carninci,P. and Hayashizaki,Y.
TITLE	High-efficiency full-length cDNA cloning
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)
PUBLISHED	10349636
REFERENCE	2
AUTHORS	Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)
PUBLISHED	11042159
REFERENCE	3
AUTHORS	Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsuunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Onoda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,Y., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)
PUBLISHED	11076861
REFERENCE	4
AUTHORS	Kawai,J., Shingawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y., Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S., Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I., Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R., Kadofa,K., Matsuda,H., Ahbner,M., Batalov,S., Casavant,T., Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H., Kuehl,P., Lewis,S., Matsuo,Y., Nikaide,I., Pesole,G., Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M., Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H., Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N., Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C., Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D., Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P., Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P., Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H., Sato,K., Schonbach,C., Seva,T., Shibata,Y., Storch,K.F., Suzuki,H., Toyooka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L., Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawai,H., Kohseuki,S. and Hayashizaki,Y.
TITLE	RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium
JOURNAL	Functional annotation of a full-length mouse cDNA collection
PUBLISHED	Nature 409 (6821), 685-690 (2001)
REFERENCE	5
AUTHORS	Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Nikaide,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Quackenbush,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S., Beisel,K.W., Blake,J.A., Brad,D., Brusic,V., Ciothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragan,T.A., Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A., Kawai,H., Kawasawa,Y., Kedierski,R.M., King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,

Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Perteau,G., Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K., Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Inotani,K., Ishii,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.	
FANTOM Consortium	Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs	of 60,770 full-length cDNAs
Nature 420 (6915), 563-573 (2002)	Nature 420 (6915), 563-573 (2002)
12466851	12466851
6	6
Carninci,P., Kasukawa,T., Katayama,S., Gough,J., Frith,M.C., Maeda,N., Oyama,R., Ravasi,T., Lenhard,B., Wells,C., Kodzius,R., Shimokawa,K., Bajic,V.B., Brenner,S.E., Batalov,S., Forrest,A.R., Zavolan,M., Davis,M.J., Wilming,L.G., Aidinis,V., Allen,J.E., Ambesi-Imbimoto,A., Apweiler,K., Aturaliya,R.N., Bailey,T.L., Bansal,M., Baxter,L., Beisel,K.W., Bersano,T., Bono,H., Chalk,A.M., Chiu,K.P., Choudhary,V., Christoffels,A., Clutterbuck,D.R., Crowe,M.L., Dalla,E., Dalrymple,B.P., de Bono,B., Della Gatta,G., di Bernardo,D., Down,T., Engstrom,P., Fagioli,M., Gatta,G., Fletcher,C.F., Fukushima,T., Furuno,M., Futaki,S., Gariboldi,M., Georgin-Hemming,P., Giegeras,T.R., Gojobori,T., Green,R.E., Gustincich,S., Harbers,M., Hayashi,Y., Hensch,T.K., Hirokawa,N., Hill,D., Hummel,L., Iacono,M., Ikeo,K., Iwama,A., Ishikawa,T., Jakt,M., Kanapin,A., Katoh,M., Kawasawa,Y., Kelso,J., Kitamura,H., Kitano,H., Kollias,G., Krishnan,S.P., Kruger,A., Kummerfeld,S.K., Kurochkin,I.V., Lareau,L.F., Lazarevic,D., Lipovich,L., Liu,J., Liuni,S., McWilliam,S., Madan Babu,M., Madera,M., Marchionni,L., Matsuda,H., Matsuzawa,S., Miki,H., Mignone,F., Miyake,S., Morris,K., Mottagui-Tabar,S., Mulder,N., Nakano,N., Nakachi,H., Ng,P., Nilsson,R., Nishiguchi,S., Nishikawa,S., Nori,F., Ohara,O., Okazaki,Y., Orlando,V., Pang,K.C., Pavan,W.J., Pavese,G., Pesole,G., Petrovsky,N., Piazza,S., Reid,J., Reid,J.F., Ring,B.Z., Ringwald,M., Rost,B., Ruan,Y., Salzberg,S.L., Sandelin,A., Schneider,C., Schonbach,C., Sekiguchi,K., Semple,C.A., Seno,S., Sessa,L., Sheng,Y., Shibata,Y., Shimada,H., Shimada,K., Silva,D., Sinclair,B., Sperling,S., Stupka,E., Sugura,K., Sultana,R., Takenaka,Y., Taki,K., Tammoja,K., Tan,S.L., Tang,S., Taylor,M.S., Tegner,J., Teichmann,S.A., Ueda,H.R., van Nimwegen,E., Verardo,R., Wei,C.L., Yagi,K., Yamanishi,H., Zabarovsky,E., Zhu,S., Zimmer,A., Hide,W., Bult,C., Grimmond,S.M., Teasdale,R.D., Liu,E.T., Brusic,V., Quackenbush,J., Wahlestedt,C., Mattick,J.S., Hume,D.A., Kai,C., Sasaki,D., Tomaru,Y., Fukuda,S., Kanamori-Katayama,M., Suzuki,M., Aoki,J., Arakawa,T., Iida,J., Imamura,K., Itoh,M., Kato,T., Kawaji,H., Kawagashira,N., Kawashima,T., Kojima,M., Kondo,S., Konno,H., Nakano,K., Ninomiya,N., Nishio,T., Okada,M., Plessey,C., Shibata,K., Shiraki,T., Suzuki,S., Tagami,M., Waki,K., Watahiki,A., Okamura-Oho,Y., Suzuki,H., Kawai,J. and Hayashizaki,Y.	
FANTOM Consortium	The transcriptional landscape of the mammalian genome
Science 309 (5740), 1559-1563 (2005)	Science 309 (5740), 1559-1563 (2005)
16141072	16141072
7	7
Katayama,S., Tomaru,Y., Kasukawa,T., Waki,K., Nakanishi,M., Nakamura,M., Nishida,H., Yap,C.C., Suzuki,M., Kawai,J., Suzuki,H., Carninci,P., Hayashizaki,Y., Wells,C., Frith,M., Ravasi,T., Pang,K.C., Hallinan,J., Mattick,J., Hume,D.A., Lipovich,L., Batalov,S., Engstrom,P.G., Mizuno,Y., Faghihi,M.A., Sandelin,A., Chalk,A.M., Mottagui-Tabar,S., Liang,Z., Lenhard,B. and Wahlestedt,C.	
RIKEN Genome Exploration Research Group	Antisense transcription in the mammalian transcriptome
Science 309 (5740), 1564-1566 (2005)	Science 309 (5740), 1564-1566 (2005)



Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Frith,M.C.,  
 Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,  
 Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,  
 Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,  
 Rogers,J., Birney,E. and Hayashizaki,Y.  
 FANTOM Consortium

CONSTRM  
 TITLE  
 Analysis of the mouse transcriptome based on functional annotation

JOURNAL  
 PUBMED  
 Of 60,770 full-length cDNAs  
 Nature 420 (6915), 563-573 (2002)

REFERENCE  
 12466851

AUTHORS

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C.,  
 Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., Kodzius, R.,  
 Shimokawa, K., Bajic, V.B., Brenner, S.E., Batalov, S., Forrest, A.R.,  
 Zavolan, M., Davis, M.J., Wilming, L.G., Aidinis, V., Allen, J.E.,  
 Ambesi-Impombato, A., Apweiler, R., Aturaliya, R.N., Bailey, T.L.,  
 Banaal, M., Baxter, L., Beisel, K.W., Bersano, T., Bono, H., Chalk, A.M.,  
 Chiu, K.P., Choudhary, V., Christoffels, A., Clutterbuck, D.R.,  
 Crowe, M.L., Dalla, E., Dalrymple, B.P., de Bono, B., Della Gatta, G.,  
 di Bernardo, D., Down, T., Engstrom, P., Fagioli, M., Faulkner, G.,  
 Fletcher, C.F., Fukushima, T., Furuno, M., Futaki, S., Gariboldi, M.,  
 Georgii-Hemming, P., Gingeras, T.R., Gojobori, T., Green, R.E.,  
 Gustincich, S., Harbers, M., Hayashi, Y., Hensch, T.K., Hirokawa, N.,  
 Hill, D., Humineci, L., Iacono, M., Ikeo, K., Iwama, A., Ishikawa, T.,  
 Jakt, M., Kanapin, A., Katoh, M., Kawasawa, Y., Kelso, J., Kitamura, H.,  
 Kitano, H., Kollias, G., Krishnan, S.P., Kruger, A., Kummerfeld, S.K.,  
 Kurochkin, I.V., Lareau, L.F., Lazarevic, D., Lipovich, L., Liu, J.,  
 Liuni, S., McWilliam, S., Madan, Babu, M., Madera, M., Marchionni, L.,  
 Matsuda, H., Matsuzawa, S., Miki, H., Mignone, F., Miyake, S.,  
 Morita, K., Mottagui-Tabar, S., Mulder, N., Nakano, N., Nakachi, H.,  
 Ng, P., Nilsson, R., Nishiguchi, S., Nishikawa, S., Nori, F., Ohara, O.,  
 Okazaki, Y., Orlando, V., Pang, K.C., Pavan, W.J., Pavese, G.,  
 Pesole, G., Petrovsky, N., Piazza, S., Reed, J., Reid, J.F., Ring, B.Z.,  
 Ringwald, M., Rost, B., Ruan, Y., Salzberg, S.L., Sandelin, A.,  
 Schneider, C., Schonbach, C., Sekiguchi, K., Semple, C.A., Seno, S.,  
 Sessa, L., Sheng, Y., Shibata, Y., Shimada, H., Shimada, K., Silva, D.,  
 Sinclaire, B., Sperling, S., Stupka, E., Sugura, K., Sultana, R.,  
 Takenaka, Y., Taki, K., Tammoja, K., Tan, S.L., Tang, S., Taylor, M.S.,  
 Tegner, J., Teichmann, S.A., Ueda, H.R., van Nimwegen, E., Verardo, R.,  
 Wei, C.L., Yagi, K., Yamashita, H., Zabarovsky, E., Zhu, S., Zimmer, A.,  
 Hide, W., Bult, C., Grimmond, S.M., Teasdale, R.D., Liu, E.T.,  
 Brusic, V., Quackenbush, J., Wahlestedt, C., Mattick, J.S., Hume, D.A.,  
 Kai, C., Sasaki, D., Tomaru, Y., Fukuda, S., Kanamori-Katayama, M.,  
 Suzuki, M., Aoki, J., Arakawa, T., Iida, J., Imamura, K., Itoh, M.,  
 Kato, T., Kawaji, H., Kawagashira, N., Kawashima, T., Kojima, M.,  
 Kondo, S., Konno, H., Nakano, K., Ninomiya, N., Nishio, T., Okada, M.,  
 Plessey, C., Shibata, K., Shiraki, T., Suzuki, S., Tagami, M., Waki, K.,  
 Watahiki, A., Okamura-Oho, Y., Suzuki, H., Kawai, J. and Hayashizaki, Y.  
 FANTOM Consortium

CONSTRM  
 TITLE  
 The transcriptional landscape of the mammalian genome

JOURNAL  
 PUBMED  
 Science 309 (5740), 1559-1563 (2005)

REFERENCE  
 16141072

AUTHORS

Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M.,  
 Nakamura, M., Nishida, H., Yap, C.C., Suzuki, M., Kawai, J., Suzuki, H.,  
 Carninci, P., Hayashizaki, Y., Wells, C., Frith, M., Ravasi, T.,  
 Pang, K.C., Hallinan, J., Mattick, J., Hume, D.A., Lipovich, L.,  
 Batalov, S., Engstrom, P.G., Mizuno, Y., Faghihi, M.A., Sandelin, A.,  
 Chalk, A.M., Mottagui-Tabar, S., Liang, Z., Lenhard, B. and  
 Wahlestedt, C.

CONSTRM  
 TITLE  
 RIKEN Genome Exploration Research Group

JOURNAL  
 PUBMED  
 Antisense transcription in the mammalian transcriptome

Science 309 (5740), 1564-1566 (2005)

REFERENCE  
 16141073

AUTHORS

Arakawa, T., Carninci, P., Fukuda, S., Hashizume, W., Hayashida, K.,  
 Hori, F., Iida, J., Imamura, K., Imotani, K., Itoh, M., Kanagawa, S.,  
 Kawai, J., Kojima, H., Konno, H., Murata, M., Nakamura, M., Ninomiya, N.,  
 Nishiyori, K., Nomura, K., Ohno, M., Sakazume, N., Sano, H., Sasaki, D.,  
 Shibata, K., Shiraki, T., Tagami, M., Tagami, Y., Waki, K., Watahiki, A.,  
 Muramatsu, M. and Hayashizaki, Y.

CONSTRM  
 TITLE  
 Direct Submission

JOURNAL  
 PUBMED  
 Submitted (14-APR-2004) Yoshihide Hayashizaki, The Institute of

Physical and Chemical Research (RIKEN), Laboratory for Genome  
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
 RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
 Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp,  
 URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,  
 Fax: 81-45-503-9216)

CDNA library was prepared and sequenced in Mouse Genome  
 Encyclopedia Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
 Division of Experimental Animal Research in Riken contributed to  
 prepare mouse tissues.

Tissues were provided by Dr. John Todd (Dept. of Medical Genetics  
 Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome  
 Trust/MRC building Addenbrookes Hospital Cambridge) whose  
 assistance we gratefully acknowledge.

Please visit our web site for further details.

URL: http://genome.gsc.riken.jp/

URL: http://fantom.gsc.riken.jp/

Location/Qualifiers

source

1. .2923  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="NOD"  
 /db\_xref="FANTOM DB:F830206103"  
 /db\_xref="taxon:10090"  
 /clone="F830206103"  
 /tissue\_type="activated spleen"  
 /cruise\_lib="RIKEN full-length enriched mouse cDNA library"  
 290. .610

CDS

/notes="unnamed protein product; Transformer-2 protein  
 homolog [Mus musculus] (UniProt Q8BVX0, evidence: FASTY,  
 94.4%ID, 100%length, match=318)  
 putative"  
 /codon\_start=1  
 /protein\_id="BAB42995.1"  
 /db\_xref="GI:74192357"  
 /translation="MAGSGPCVPCPSQPSFLAADWYLAWSYPLLSPLSLASHLCS  
 EVRVGAVGQEAAGSLRPSDGRGFLRWLGGPPPLLRPSQRLGVRSGCCPGGES  
 LARD"

ORIGIN

Query Match 89.5%; Score 17; DB 6; Length 2923;  
 Best Local Similarity 100.0%; Pred. No. 3.7e+03;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TTACTTCATAGTCCTTT 18

|||||

Db 1677 TTACTTCATAGTCCTTT 1661

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)

JOURNAL  
PUBMED  
REFERENCE  
AUTHORS

3  
Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P.,  
Konno, H., Akiyama, J., Nishi, K., Kitsuunai, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,  
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)

TITLE  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS

4  
The RIKEN Genome Exploration Research Group Phase II Team and the  
FANTOM Consortium.

TITLE  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS

Functional annotation of a full-length mouse cDNA collection  
Nature 409, 685-690 (2001)

5  
The FANTOM Consortium, the RIKEN Genome Exploration Research Group  
Phase I and II Team.

TITLE  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS

Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)

TITLE  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS

6  
RIKEN Genome Exploration Research Group, Genome Science Group  
(Genome Network Core Team) and the FANTOM Consortium.  
Antisense Transcription in the Mammalian Transcriptome  
Science 309, 1564-1566 (2005)

TITLE  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS

7  
The FANTOM Consortium, Riken Genome Exploration Research Group and  
Genome Science Group (Genome Network Project Core Group).  
The Transcriptional Landscape of the Mammalian Genome  
Science 309, 1559-1563 (2005)

TITLE  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS

8 (bases 1 to 2942)  
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,  
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,  
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,  
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,  
Koya, S., Kurihara, C., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,  
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,  
Okazaki, Y., Saichoh, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,  
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,  
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,  
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,  
Muramatsu, M., and Hayashizaki, Y.

TITLE  
JOURNAL

Direct Submission  
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa, 230-0045, Japan (E-mail: genome-res@gsc.riken.jp,  
URL: http://genome-gsc.riken.jp/, Tel: 81-45-503-9222,  
Fax: 81-45-503-9216)

COMMENT

cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.

Please visit our web site for further details.  
URL: http://genome-gsc.riken.jp/  
URL: http://fantom-gsc.riken.jp/.

FEATURES  
source

Location/Qualifiers  
1. .2942  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM\_DB:4921532C16"  
/db\_xref="taxon:10090"  
/clone="4921532C16"

/sex="male"  
/tissue\_type="testis"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="adult"  
1. .2942  
/note="putative  
unclassifiable"

misc\_feature

ORIGIN

Query Match 89.5%; Score 17; DB 6; Length 2942;  
Best Local Similarity 100.0%; Pred. No. 3.7e+03;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 TTTACTTTCATAGTCTTT 18  
|||||

Db 1707 TTTACTTTCATAGTCTTT 1691  
|||||

RESULT 91

AKI70599/c  
LOCUS AKI70599 3785 bp mRNA linear HTC 21-SEP-2005  
DEFINITION Mus musculus NOD-derived CD11c +ve dendritic cells cDNA, RIKEN  
full-length enriched library, clone:F630109013  
product:Transformer-2 protein homolog [Mus musculus], full insert  
sequence.  
AKI70599  
VERSION AKI70599.1 GI:74215396  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muroidae; Muridae; Murinae; Mus.

REFERENCE

1 Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)  
10349636  
2  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)  
11042159  
3  
Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P.,  
Konno, H., Akiyama, J., Nishi, K., Kitsuunai, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,  
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
11076861

REFERENCE

4  
Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,  
Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,  
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I.,  
Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,  
Kadoya, K., Matsuura, H., Ashburner, M., Batalov, S., Casavant, T.,  
Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,  
Kuehl, P., Lewis, S., Matsuo, Y., Nikolaio, I., Pesole, G.,  
Quackenbush, J., Schriml, L. M., Staib, F., Suzuki, R., Tonita, M.,  
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,  
Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,  
Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C.,  
Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,  
Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P.,  
Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P.,  
Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,

Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H., Toyo-oka, K., Wang, K.H., Weitz, C., Whittaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S. and Hayashizaki, Y.

RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium

Functional annotation of a full-length mouse cDNA collection

Nature 409 (6821), 685-690 (2001)

11217851

5

CONSRMTM

JOURNAL

PUBMED

REFERENCE

AUTHORS

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Mikado, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V., Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierzki, R.M., King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shingawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.

FANTOM Consortium

Analysis of the mouse transcriptome based on functional annotation

of 60,770 full-length cDNAs

Nature 420 (6915), 563-573 (2002)

12466851

6

CONSRMTM

JOURNAL

PUBMED

REFERENCE

AUTHORS

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., Kodzius, R., Shimokawa, K., Bajic, V.B., Brenner, S.E., Batalov, S., Forrest, A.R., Zavolan, M., Davis, M.J., Wilming, L.G., Aidinis, V., Allen, J.E., Ambesi-Impombato, A., Apweiler, R., Aturaliya, R.N., Bailey, T.L., Bansal, M., Baxter, L., Belsel, K.W., Bersano, T., Bono, H., Chalk, A.M., Chiu, K.P., Choudhary, V., Christoffels, A., Clutterbuck, D.R., Crowe, M.L., Dalla, E., Dalrymple, B.P., de Bono, B., Della Gatta, G., di Bernardo, D., Down, T., Engstrom, P., Fagioli, M., Faulkner, G., Fletcher, C.F., Fukushima, T., Furuno, M., Futaki, S., Gariboldi, M., Georgii-Hemming, P., Gingeras, T.R., Gojobori, T., Green, R.E., Gustincich, S., Harbers, M., Hayashizaki, Y., Hensch, T.K., Hirokawa, N., Hill, D., Humaneck, L., Iacono, M., Ikeo, K., Iwama, A., Ishikawa, T., Jakt, M., Kanapin, A., Katoh, M., Kawasawa, Y., Kelso, J., Kitamura, H., Kitano, H., Kollias, G., Krishnan, S.P., Kruger, A., Kummerfeld, S.K., Kurochkin, I.V., Lareau, L.F., Lazarevic, D., Lipovich, L., Liu, J., Liuni, S., McWilliam, S., Madan, Babu, M., Madera, M., Marchionni, L., Matsuda, H., Matsuzawa, S., Miki, H., Mignone, F., Miyake, S., Moris, K., Mottagui-Tabar, S., Mulder, N., Nakano, N., Nakachi, H., Ng, P., Nilsson, R., Nishiguchi, S., Nishikawa, S., Nori, F., Ohara, O., Okazaki, Y., Orlando, V., Pang, K.C., Pavan, W.J., Pavesi, G., Pesole, G., Petrovsky, N., Piazza, S., Reed, J., Reid, J.F., Ring, B.Z., Ringwald, M., Rost, B., Ruan, Y., Salzberg, S.L., Sandelin, A., Schneider, C., Schonbach, C., Sekiguchi, K., Semple, C.A., Seno, S., Sessa, L., Sheng, Y., Shibata, Y., Shimada, K., Shimada, K., Silva, D., Sincclair, B., Sperling, S., Stupka, E., Sugiyara, K., Sultana, R., Takenaka, Y., Taki, K., Tammoja, K., Tan, S.L., Tang, S., Taylor, M.S., Tegner, J., Teichmann, S.A., Ueda, H.R., van Nimwegen, E., Vervardo, R., Wei, C.L., Yagi, K., Yamanishi, H., Zabarovsky, E., Zhu, S., Zimmer, A., Hide, W., Bult, C., Grimmond, S.M., Teasdale, R.D., Liu, E.T.,

Brusic, V., Quackenbush, J., Wahlestedt, C., Mattick, J.S., Hume, D.A., Kai, C., Sasaki, D., Tomaru, Y., Fukuda, S., Kanamori-Katayama, M., Suzuki, M., Aoki, J., Arakawa, T., Iida, J., Imamura, K., Itoh, M., Kato, T., Kawaji, H., Kawagashira, N., Kawashima, T., Kojima, M., Kondo, S., Konno, H., Nakano, K., Ninomiya, N., Nishio, T., Okada, M., Plessey, C., Shibata, K., Shiraki, T., Suzuki, S., Tagami, M., Waki, K., Watanabe, A., Okamura-Oho, Y., Suzuki, H., Kawai, J. and Hayashizaki, Y.

The transcriptional landscape of the mammalian genome

Science 309 (5740), 1559-1563 (2005)

16141072

7

CONSRMTM

JOURNAL

PUBMED

REFERENCE

AUTHORS

Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yap, C.C., Suzuki, M., Kawai, J., Suzuki, H., Carninci, P., Hayashizaki, Y., Wells, C., Frith, M., Ravasi, T., Pang, K.C., Hallinan, J., Mattick, J., Hume, D.A., Lipovich, L., Batalov, S., Engstrom, P.G., Mizuno, Y., Faghihi, M.A., Sandelin, A., Chalk, A.M., Mottagui-Tabar, S., Liang, Z., Lenhard, B. and Wahlestedt, C.

RIKEN Genome Exploration Research Group

Antisense transcription in the mammalian transcriptome

Science 309 (5740), 1564-1566 (2005)

16141073

8

(bases 1 to 3785)

Arakawa, T., Carninci, P., Fukuda, S., Hashizume, W., Hayashida, K., Hori, F., Iida, J., Imamura, K., Imotani, K., Itoh, M., Kanegawa, S., Kawai, J., Kojima, M., Konno, H., Murata, M., Nakamura, M., Ninomiya, N., Nishiyori, H., Nomura, K., Ohno, M., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shiraki, T., Tagami, M., Tagami, Y., Waki, K., Watanabe, A., Muramatsu, M. and Hayashizaki, Y.

Direct Submission

Submitted (14-APR-2004) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gs.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge.

Please visit our web site for further details.

URL: http://genome.gsc.riken.jp/

URL: http://fantom.gsc.riken.jp/

Location/Qualifiers

1. .3785

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="NOD"

/db\_xref="FANTOM,DB:F630109013"

/db\_xref="taxon:10090"

/clone="F630109013"

/cell\_type="NOD-derived CD11c +ve dendritic cells"

/clone\_lib="RIKEN full-length enriched mouse cDNA library"

329. .652

/notes="unnamed protein product; Transformer-2 protein homolog [Mus musculus] (Uniprot|Q8BVK0, evidence: FASTA, 95.3%ID, 100%length, match=321)

/putative"

/codon\_start=1

/protein\_id="BAE41904.1"

/db\_xref="GI:74215397"

/translation="MAGSGPCVPCPTSPORSFLADWYLAWSYPLLSFLSASHLC SEVRVPGVQAEAGRSRLPDSGCRFLRWHLGPPPLLRSGRLGVRSGCCPGGE SLARD"

FEATURES

source

CDS

ORIGIN



Query Match 89.5%; Score 17; DB 6; Length 3785;  
 Best Local Similarity 100.0%; Pred. No. 3.8e+03;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TTTACTTCATAGTCTTT 18  
 |||||  
 Db 1719 TTTACTTCATAGTCTTT 1703

RESULT 92  
 AK076230/c

LOCUS  
 DEFINITION AK076230 3796 bp mRNA linear HTC 02-SEP-2005  
 Mus musculus 1 day pregnant adult female mammary gland cDNA, RIKEN  
 full-length enriched library, clone:4222403J09 product:inferred:  
 RIKEN cDNA 1500010G04 gene, full insert sequence.

ACCESSION AK076230  
 VERSION AK076230.1 GI:26345227  
 KEYWORDS HTC; CAP trapper.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE 1  
 Carninci,P. and Hayashizaki,Y.  
 High-efficiency full-length cDNA cloning  
 Meth. Enzymol. 303, 19-44 (1999)  
 PUBMED 10349636

REFERENCE 2  
 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,  
 Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new genes  
 Genome Res. 10 (10), 1617-1630 (2000)  
 JOURNAL 11042159  
 PUBMED 11042159

REFERENCE 3  
 Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,  
 Konno,H., Akiyama,J., Nishi,K., Kitasumi,T., Tashiro,H., Itoh,M.,  
 Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,  
 Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,  
 Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,  
 Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsumura,S., Kawai,J.,  
 Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multicapillary sequencer  
 Genome Res. 10 (11), 1757-1771 (2000)  
 JOURNAL 11076861  
 PUBMED 11076861

REFERENCE 4  
 The RIKEN Genome Exploration Research Group Phase II Team and the  
 FANTOM Consortium.  
 Functional annotation of a full-length mouse cDNA collection  
 Nature 409, 685-690 (2001)

REFERENCE 5  
 The FANTOM Consortium, the RIKEN Genome Exploration Research Group  
 Phase I and II Team.  
 Analysis of the mouse transcriptome based on functional annotation  
 of 60,770 full-length cDNAs  
 Nature 420, 563-573 (2002)

REFERENCE 6  
 RIKEN Genome Exploration Research Group, Genome Science Group  
 (Genome Network Core Team) and the FANTOM Consortium.  
 Antisense Transcription in the Mammalian Transcriptome  
 Science 309, 1564-1566 (2005)

REFERENCE 7  
 The FANTOM Consortium, Riken Genome Exploration Research Group and  
 Genome Science Group (Genome Network Project Core Group).  
 The Transcriptional Landscape of the Mammalian Genome  
 Science 309, 1559-1563 (2005)

REFERENCE 8  
 (bases 1 to 3796)  
 Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,  
 Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,  
 Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,  
 Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,

Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,  
 Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,  
 Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,  
 Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,  
 Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,  
 Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,  
 Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,  
 Muramatsu,M. and Hayashizaki,Y.  
 Direct Submission  
 Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of  
 Physical and Chemical Research (RIKEN), Laboratory for Genome  
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
 RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
 Kanagawa, 230-0045, Japan (E-mail:genome-res@gsc.riken.jp,  
 URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,  
 Fax:81-45-503-9216)  
 COMMENT  
 cDNA library was prepared and sequenced in Mouse Genome  
 Encyclopedia Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
 Division of Experimental Animal Research in Riken contributed to  
 prepare mouse tissues.  
 Please visit our web site for further details.  
 URL:http://genome.gsc.riken.jp/  
 URL:http://fantom.gsc.riken.jp/  
 FEATURES  
 source  
 1. .3796  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="FANTOM DB:4222403J09"  
 /db\_xref="taxon:10090"  
 /clone="4222403J09"  
 /sex="female"  
 /tissue\_type="mammary gland"  
 /clone\_lib="RIKEN full-length enriched mouse cDNA library"  
 /dev\_stage="1 day pregnant adult"  
 333..656  
 /notes="unnamed protein product; inferred: RIKEN cDNA  
 1500010G04 gene (UniGene|Mm.196598, evidence: UG)  
 putative"  
 /codon\_start=1  
 /protein\_id="BAC36264.1"  
 /db\_xref="GI:26345228"  
 /translation="MASGSPVPCPTSPORSFLAADWYLAWSYPLLSLPFLASHLC  
 SEVRPGAVGQEAAGRSRLRPSDSCFLRWHLWAPPPLLRPSQRLGVKSARGCCPGGE  
 SLVRD"  
 ORIGIN  
 Query Match 89.5%; Score 17; DB 6; Length 3796;  
 Best Local Similarity 100.0%; Pred. No. 3.8e+03;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TTTACTTCATAGTCTTT 18  
 |||||  
 Db 1727 TTTACTTCATAGTCTTT 1711

RESULT 93  
 DE242715/c

LOCUS  
 DEFINITION DE242715 91 bp DNA linear GSS 06-JAN-2006  
 Trifolium pratense DNA, clone:RCG48925, genomic survey sequence.  
 DE242715  
 ACCESSION DE242715  
 VERSION DE242715.1 GI:84463411  
 KEYWORDS GSS.  
 SOURCE Trifolium pratense  
 ORGANISM Trifolium pratense  
 Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;  
 Trifolium.  
 1  
 Sato,S., Isobe,S., Asamizu,E., Nakamura,Y., Ohmido,N., Sakurai,N.,  
 Klimenko,I., Sasamoto,S., Wada,T., Watanabe,A. and Tabata,S.

```

TITLE      Comprehensive structural analysis of the genome of red clover
JOURNAL    (Trifolium pratense)
REFERENCE  2  (bases 1 to 91)
AUTHORS    Sato, S.
TITLE      Direct Submission
JOURNAL    Submitted (01-SEP-2005) Shusei Sato, Kazusa DNA Research Institute,
            The First Laboratory for Plant Gene Research; 2-6-7
            Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan
            (E-mail: ssato@kazusa.or.jp, URL: http://www.kazusa.or.jp/,
            Tel: 81-438-52-3935 (ex.2337), Fax: 81-438-52-3934)

FEATURES   Location/Qualifiers
            1..91
            /organism="Trifolium pratense"
            /mol_type="genomic DNA"
            /db_xref="taxon:57577"
            /clone="RCG48925"

ORIGIN
Query Match      86.3%; Score 16.4; DB 14; Length 91;
Best Local Similarity 94.4%; Pred. No. 4.8e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1  CTTTACTTCATGCTCTTT 18
        |||||
Db      29  CTTTACTTCATGCTCTTT 12

RESULT 94
BH497760/c
LOCUS      BH497760.1 GI:17705864
DEFINITION      172 bp DNA linear GSS 13-DEC-2001
                BOHPC14TF BOHP Brassica oleracea genomic clone BOHPC14, genomic
                survey sequence.
ACCESSION      BH497760
VERSION        BH497760.1 GI:17705864
KEYWORDS
SOURCE
ORGANISM      Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 172)
Ayele,M., Haas,B.J., Kumar,N., Wu,H., Xiao,Y., Van Aken,S.,
Utterback,T.R., Wortman,J.R., White,O.R. and Town,C.D.
Whole genome shotgun sequencing of Brassica oleracea and its
application to gene discovery and annotation in Arabidopsis
Genome Res. 15 (4), 487-495 (2005)
15805490
PUBMED
Other_GSSs: BOHPC14TR
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.

FEATURES   Location/Qualifiers
            1..172
            /organism="Brassica oleracea"
            /mol_type="genomic DNA"
            /strain="TO1000DH3"
            /db_xref="taxon:3712"
            /clone="BOHPC14"
            /clone_lib="BOHP"
            /note="Vector: PHOS1; Site_1: BstXI; 2-3 kb sheared
            genomic DNA inserted into PHOS1 using BstXI linkers"

ORIGIN
Query Match      86.3%; Score 16.4; DB 11; Length 172;
Best Local Similarity 94.4%; Pred. No. 5.2e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      2  TTTACTTCATAGTCTTTG 19
        |||||
Db      150 TTTACTTCATAGTCTTTG 133

BB419641      205 bp mRNA linear EST 16-JUL-2000
BB419641      RIKEN full-length enriched, 12 days embryo spinal cord Mus
musculus cDNA clone C530017P24 3' similar to Y12883 M.musculus mRNA
for nonseelenium glutathione peroxidase, mRNA sequence.
BB419641
EST.
BB419641.1 GI:9240996
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.
1 (bases 1 to 205)
Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T.,
Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C.,
Kusakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H.,
Okazaki,Y., Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K.,
Shibata,K., Shibata,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Tagawa,A.,
Takahashi,F., Tomimaga,N., Toya,T., Tsunoda,Y., Watahiki,A.,
Watanabe,S., Yamamura,T., Yamataka,I., Yano,R., Yasunishi,A.,
Yokota,T., Yoshida,K., Yoshiki,A., Yoshino,M., Muramatsu,M. and
Hayashizaki,Y.
RIKEN Mouse ESTs (Konno,H., et al.)
Unpublished (2000)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoaka,S.,
Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermotabilization and thermostaotivation of thermolabile enzymes by
trehalose and its application for the synthesis of full length
cDNA. Proc Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh,M., Kitsuai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M.,
Okazaki,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.
            Location/Qualifiers
            1..205
            /organism="Mus musculus"
            /mol_type="mRNA"
            /db_xref="taxon:10090"
            /clone="C530017P24"
            /tissue type="spinal cord"
            /dev_stage="12 days embryo"
            /lab_host="DH10B"
            /clone_lib="RIKEN full-length enriched, 12 days embryo
            spinal cord"
            /note="Site 1: Sali; Site 2: BamHI; cDNA library was
            prepared and sequenced in Mouse Genome Encyclopedia
            Project of Genome Exploration Research Group in Riken

```



Sciences Center (OSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel.: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.jp, URL: <http://genome.gsc.riken.jp/>  
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S.,  
Sasaki, N., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.  
Thermotabilization and thermoinactivation of thermostable enzymes by  
trehalose and its application for the synthesis of full length  
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
Itoh, M., Kitsumai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,  
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M.,  
Okazaki, Y., and Hayashizaki, Y.  
Automated filtration-based high-throughput plasmid preparation  
system. Genome Res. 9 (5), 463-470 (1999)  
Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,  
19-44 (1999)  
Please visit our web site (<http://genome.rtc.riken.go.jp>) for  
further details.

FEATURES  
source

```
Location/Qualifiers
1. .279
/or/ganism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone_id="A23060K02"
/sex="male"
/tissue_type="hypothalamus"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, adult male hypothalamus"
/note="Site1: Salt; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAGATCCACGAGCTCTTTTTTTTTTTTNN 3'], cDNA was prepared by using trihalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTAATTAATATCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda phage."
PIC "
```

## ORIGIN

Query Match	86.3%	Score 16.4;	DB 7;	Length 279;
Best Local Similarity	94.4%;	Pred. No. 5.5e+03;		
Matches 17;	Conservative	0;	Mismatches 1;	Indels 0;
QY	2	TTTACTTCATAGTCTTTG	19	
Db	156	TTTCTTCATAGTCTTTG	173	

RESULT 98

[illegible]

REFERENCE  
AUTHORS

Sciurognathi; Muroidea; Muridae; Murinae; Mus.

1 (Bases 1 to 290)

Konno H., Aizawa K., Akahira, S., Akiyama, J., Arakawa, T., Hayatsu, N., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurlihar, C., Kiyosabe, M., Matsumaya, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Saito, C., Sato, K., Shibata, K., Shibata, Y., Shigenoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomingaga, N., Toyota, T., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yamanaka, I., Yano, K., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Havashizaki, Y.

TITLE	JOURNAL	COMMENT
1. The Role of the Teacher in the Classroom	Journal of Educational Research	1980, Vol. 83, No. 1, pp. 1-10
2. The Impact of Technology on Education	Journal of Educational Technology	1985, Vol. 10, No. 2, pp. 15-25
3. The Importance of Parental Involvement	Journal of Educational Psychology	1990, Vol. 82, No. 3, pp. 201-210
4. The Effect of Student Motivation on Learning	Journal of Educational Research	1995, Vol. 98, No. 4, pp. 301-310
5. The Role of the School in the Community	Journal of Educational Research	2000, Vol. 103, No. 1, pp. 1-10

RIKEN Mouse ESTs (Konno, H., et al.)  
Unpublished (2000)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Sushiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.jp, URL: <http://genome.gsc.riken.jp/>  
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S.,  
Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.:  
Thermotranscription and thermotranscription of thermotranscription enzymes by  
trichostatin A and its application for the synthesis of full length  
cDNA. *Proc. Natl. Acad. Sci. U.S.A.* 95 (2), 520-524 (1998)  
Itoh, M., Kiteunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,  
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M.,  
Okazaki, Y. and Hayashizaki, Y.:  
Automated filtration-based high-throughput plasmid preparation  
system. *Genome Res.* 9 (5), 463-470 (1999)  
Carninci, P. and Hayashizaki, Y.:  
High-efficiency full-length cDNA cloning. *Methods Enzymol.* 303,  
19-44 (1999)  
Please visit our web site (<http://genome.rtc.riken.go.jp>) for  
further details.

## FEATURES

**Location/Qualifiers**

1. .290

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="5330437J01"

/sex="male"

/tissue\_type="pituitary gland"

/dev\_stage="adult"

/lab\_host="DH10B"

/clone\_lib="RIKEN full-length enriched, adult male pituitary gland"

/note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAAGATTCGAAGCTCTTTTCTTTTTTTTN 3'], cDNA was prepared by using thermostable thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 3.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAATTCGAGTAATTAAATTAATCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified plucscript KS(+) after bulk excision from Lambda ECL1 cloning site 5' end,SalI, 3' end: BamI."

## ORIGIN

Query Match 86.3%; Score 16.4; DB 7; Length 290;

Best Local Similarity 94.4%; Pred. No. 5.5e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TTTTCTTCATAGTCTTTG 19  
Db 166 TTTTCTTCATAGTCTTTG 183

RESULT 99  
CZ443147 293 bp mRNA linear GSS 08-APR-2005  
LOCUS CMHD-GT\_211D5-3 GTL\_R1\_pGTIV1 Mus musculus cDNA clone  
DEFINITION CMHD-GT\_211D5-3 3', mRNA sequence.

ACCESSION CZ443147  
VERSION  
KEYWORDS GSS.  
SOURCE CMHD43147.1 GI:62426451

ORGANISM Mus musculus (house mouse)

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridea; Muridae; Murinae; Mus.  
1 (bases 1 to 293)

AUTHORS Stanford, W.L.

TITLE www.cmhd.ca

JOURNAL Unpublished (2002)

COMMENT Contact: Stanford WL

Institute of Biomaterials & Biomedical Engineering

University of Toronto

407 Rosebrugh Bldg., 4 Taddle Creek Rd., Toronto, Ontario, Canada  
M5S 3G9

Tel: 416 946 8379

Fax: 416 978 4317

Email: william.stanford@utoronto.ca

pGTIV1 Gene trap insertion. The sequence tag is generated by 3'  
race. The ES cell line harboring this insertion of the target gene  
is available through the following web site:  
[http://pokey.ibe.utoronto.ca/sequence\\_report.php?id=211D5](http://pokey.ibe.utoronto.ca/sequence_report.php?id=211D5).

Class: Gene Trap.

FEATURES Location/Qualifiers

1..293  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="129S3"  
/db\_xref="taxon:10090"  
/clone="CMHD-GT\_211D5-3"  
/sex="Male"  
/cell\_type="Embryonic Stem Cell"  
/cell\_line="R1"  
/clone\_lib="GTL\_R1\_pGTIV1"  
/note="Vector: pGTIV1"

ORIGIN

Query Match 86.3%; Score 16.4; DB 13; Length 293;  
Best Local Similarity 94.4%; Pred. No. 5.5e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTT 18  
Db 99 CTTTACTTCATAGTCATT 116

RESULT 100  
DN761057/c 300 bp mRNA linear EST 01-JAN-2006  
LOCUS DN761057  
DEFINITION G.hir-stem 3564 PCR amplified cotton stem cDNA Gossypium hirsutum  
cDNA, mRNA sequence.

ACCESSION DN761057  
VERSION  
KEYWORDS DN761057.1 GI:84146815  
SOURCE EST.

ORGANISM Gossypium hirsutum (upland cotton)

Gossypium hirsutum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;

rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.  
1 (bases 1 to 300)

AUTHORS Taliercio, E.

TITLE Cotton ESTs from Stems

JOURNAL Unpublished (2005)

COMMENT Contact: Taliercio, Earl

Earl Taliercio

USDA-ARS

141 Experiment Station Rd, Stoneville, MS 38756, USA

Tel: 6626865363

Fax: 6626865218

Email: Etaliercio@msa-stoneville.ars.usda.gov.

FEATURES

source

1..300  
Location/Qualifiers

/organism="Gossypium hirsutum"

/mol\_type="mRNA"

/cultivar="desl19"

/db\_xref="taxon:3635"

/lab\_host="E. coli"

/clone\_lib="PCR amplified cotton stem cDNA"

/note="Organ: 10 week old stems; Vector: Topo TA; Stems

were harvested from field grown plants"

ORIGIN

Query Match 86.3%; Score 16.4; DB 9; Length 300;

Best Local Similarity 94.4%; Pred. No. 5.5e+03;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTT 18

Db 257 CTTTGCTTCATAGTCTTT 240

Search completed: June 10, 2006, 19:35:23

Job time : 2400.6 secs

**THIS PAGE IS BLANK**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2006, 15:21:38 ; Search time 42.5917 Seconds  
(without alignments)  
834.695 Million cell updates/sec

Title: US-10-600-816-32

Perfect score: 19

Sequence: 1 cttacttcacagctcttg 19

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1403666 seqs, 935554401 residues

Total number of hits satisfying chosen parameters: 2807332

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

Database :

Issued Patents NA:\*

- 1: /EMC Celerra\_IDS3/ptodata/2/ina/1 COMB.seq.\*
- 2: /EMC Celerra\_IDS3/ptodata/2/ina/5 COMB.seq.\*
- 3: /EMC Celerra\_IDS3/ptodata/2/ina/6A COMB.seq.\*
- 4: /EMC Celerra\_IDS3/ptodata/2/ina/6B COMB.seq.\*
- 5: /EMC Celerra\_IDS3/ptodata/2/ina/7 COMB.seq.\*
- 6: /EMC Celerra\_IDS3/ptodata/2/ina/H COMB.seq.\*
- 7: /EMC Celerra\_IDS3/ptodata/2/ina/PTUS COMB.seq.\*
- 8: /EMC Celerra\_IDS3/ptodata/2/ina/PP COMB.seq.\*
- 9: /EMC Celerra\_IDS3/ptodata/2/ina/RE COMB.seq.\*
- 10: /EMC Celerra\_IDS3/ptodata/2/ina/backfiles.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	19	100.0	601	3	US-09-949-016-29173, A
C 2	19	100.0	601	3	US-09-949-016-126957
C 3	19	100.0	1212	3	US-09-188-930-249
C 4	19	100.0	1212	3	US-09-312-283C-249
C 5	19	100.0	1228	3	US-09-495-050A-223
C 6	19	100.0	2289	3	US-09-949-016-3613
C 7	19	100.0	2456	3	US-09-949-016-572
C 8	19	100.0	26086	3	US-09-949-016-15355
C 9	19	100.0	26238	3	US-09-949-016-12314
C 10	17.4	91.6	601	3	US-09-949-016-173650
C 11	17.4	91.6	50950	3	US-09-949-016-16659
C 12	16.4	86.3	407	3	US-09-513-999C-32761
C 13	16.4	86.3	15131	3	US-09-949-016-16805
C 14	16.4	86.3	22614	3	US-09-949-016-12148
C 15	16.4	86.3	22615	3	US-09-949-016-16735
C 16	16.4	86.3	107937	3	US-09-949-016-17192
C 17	16.4	86.3	139049	3	US-09-949-016-17030
C 18	16	84.2	2363	3	US-09-949-016-2159
C 19	16	84.2	22266	3	US-09-949-016-13901
C 20	16	84.2	130563	3	US-09-949-016-12273
C 21	16	84.2	131379	3	US-09-949-016-16050
C 22	15.8	83.2	601	3	US-09-949-016-48463
C 23	15.8	83.2	601	3	US-09-949-016-48464

Sequence 48465, A  
Sequence 48466, A  
Sequence 48467, A  
Sequence 48468, A  
Sequence 121554,  
Sequence 12078, A  
Sequence 1023, Ap  
Sequence 15279, A  
Sequence 15133, A  
Sequence 15990, A  
Sequence 13175, A  
Sequence 12009, A  
Sequence 17053, A  
Sequence 16435, A  
Sequence 154, App  
Sequence 141, App  
Sequence 21055, A  
Sequence 21056, A  
Sequence 21057, A  
Sequence 21058, A  
Sequence 60155, A  
Sequence 77977, A  
Sequence 77978, A  
Sequence 77979, A  
Sequence 77980, A  
Sequence 144809,  
Sequence 162942,  
Sequence 8840, Ap  
Sequence 382, App  
Sequence 659, App  
Sequence 295, App  
Sequence 295, App  
Sequence 5, Appli  
Sequence 5, Appli  
Sequence 5, Appli  
Sequence 5, Appli  
Sequence 6, Appli  
Sequence 30, Appli  
Sequence 41, Appli  
Sequence 38, Appli  
Sequence 1289, Ap  
Sequence 2698, Ap  
Sequence 6, Appli  
Sequence 16, Appli  
Sequence 16, Appli  
Patent No. 5455158  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 135, App  
Sequence 2899, Ap  
Sequence 16147, A  
Sequence 11904, A  
Sequence 14016, A  
Sequence 15836, A  
Sequence 13674, A  
Sequence 13242, A  
Sequence 574, App  
Sequence 802, App  
Sequence 16480, A  
Sequence 11868, A  
Sequence 14588, A  
Sequence 14589, A  
Sequence 17119, A  
Sequence 13498, A  
Sequence 14543, A  
Sequence 16297, A  
Sequence 12557, A  
Sequence 211, App  
Sequence 1, Appli  
Sequence 3, Appli  
Sequence 3, Appli  
US-09-949-016-48465  
US-09-949-016-48466  
US-09-949-016-48467  
US-09-949-016-48468  
US-09-949-016-121554  
US-09-949-016-12078  
US-09-620-312D-1023  
US-09-270-767-12078  
US-09-949-016-15133  
US-09-949-016-15990  
US-09-949-016-13175  
US-09-949-016-12009  
US-09-949-016-17053  
US-09-949-016-16435  
US-09-248-796A-154  
US-09-397-787-141  
US-09-949-016-21055  
US-09-949-016-21056  
US-09-949-016-21057  
US-09-949-016-21058  
US-09-949-016-60155  
US-09-949-016-77977  
US-09-949-016-77978  
US-09-949-016-77979  
US-09-949-016-77980  
US-09-949-016-144809  
US-09-949-016-162942  
US-09-949-002-8840  
US-09-614-221A-382  
US-10-104-047-659  
US-08-956-171E-295  
US-08-781-986A-295  
US-09-248-528-5  
US-09-549-108-5  
US-09-549-111-5  
US-09-549-106-5  
US-09-550-394-5  
US-09-354-147C-6  
US-08-927-939A-30  
US-09-354-147C-41  
US-09-220-132-38  
US-09-023-655-1289  
US-09-543-679A-2698  
PCT-US95-09819-6  
US-08-259-569-16  
US-08-826-885-16  
5455158-2  
US-08-551-356-1  
PCT-US93-12687-1  
US-09-566-921-135  
US-09-543-679A-2699  
US-09-949-016-16147  
US-09-949-016-11904  
US-09-949-016-14016  
US-09-949-016-15836  
US-09-949-016-12674  
US-09-949-016-13242  
US-09-949-002-574  
US-09-949-002-802  
US-09-949-016-16480  
US-09-949-016-11868  
US-09-949-016-14598  
US-09-949-016-14589  
US-09-949-016-17119  
US-09-949-016-13498  
US-09-949-016-14543  
US-09-949-016-16297  
US-09-949-016-12557  
US-09-531-120-211  
US-08-916-421B-1  
US-09-692-570-1  
US-08-989-299-3  
US-09-407-427-3

C 97	15	78.9	2415	3	US-09-635-501-3	Sequence 3, Appli	C 170	14.8	77.9	3111	3	US-09-504-393-2	Sequence 2, Appli
C 98	15	78.9	2415	4	US-09-163-648-3	Sequence 3, Appli	C 171	14.8	77.9	3111	3	US-10-053-192-2	Sequence 2, Appli
C 99	15	78.9	2920	3	US-10-158-847-137	Sequence 137, App	C 172	14.8	77.9	3177	3	US-09-248-796A-5784	Sequence 5784, App
C 100	15	78.9	2920	3	US-10-158-825-137	Sequence 137, App	C 173	14.8	77.9	4146	3	US-09-614-221A-229	Sequence 229, App
C 101	15	78.9	3388	3	US-09-949-016-5413	Sequence 5413, Ap	C 174	14.8	77.9	7680	3	US-09-210-748A-3	Sequence 3, Appli
C 102	15	78.9	3386	3	US-08-989-299-1	Sequence 1, Appli	C 175	14.8	77.9	7680	3	US-09-939-581A-3	Sequence 3, Appli
C 103	15	78.9	3386	3	US-10-158-847-141	Sequence 141, App	C 176	14.8	77.9	10034	3	US-09-771-357-102	Sequence 102, App
C 104	15	78.9	3386	3	US-09-407-427-1	Sequence 1, Appli	C 177	14.8	77.9	10034	3	US-10-059-579A-102	Sequence 102, App
C 105	15	78.9	3396	3	US-09-635-501-1	Sequence 1, Appli	C 178	14.8	77.9	11411	3	US-09-949-016-13357	Sequence 13357, A
C 106	15	78.9	3396	3	US-10-158-825-141	Sequence 141, App	C 179	14.8	77.9	15402	3	US-09-949-016-12868	Sequence 12868, A
C 107	15	78.9	3396	4	US-09-163-648-1	Sequence 1, Appli	C 180	14.8	77.9	17842	3	US-09-949-016-13826	Sequence 13826, A
C 108	15	78.9	43986	3	US-09-949-016-17155	Sequence 17155, A	C 181	14.8	77.9	23849	3	US-09-949-016-13813	Sequence 13813, A
C 109	14.8	77.9	247	3	US-09-533-559-7849	Sequence 7849, Ap	C 182	14.8	77.9	24214	3	US-09-949-016-15551	Sequence 15551, A
C 110	14.8	77.9	300	4	US-09-297-648-2037	Sequence 2037, Ap	C 183	14.8	77.9	24735	3	US-09-949-016-12850	Sequence 12850, A
C 111	14.8	77.9	363	3	US-09-513-999C-35479	Sequence 35479, A	C 184	14.8	77.9	25589	3	US-09-949-002-716	Sequence 716, App
C 112	14.8	77.9	372	3	US-09-513-999C-20504	Sequence 20504, A	C 185	14.8	77.9	26011	3	US-09-949-016-17499	Sequence 17499, A
C 113	14.8	77.9	378	3	US-09-513-999C-21946	Sequence 21946, A	C 186	14.8	77.9	29973	5	US-09-984-429-262	Sequence 262, App
C 114	14.8	77.9	473	3	US-08-513-999C-11142	Sequence 11142, A	C 187	14.8	77.9	43086	3	US-09-949-016-13059	Sequence 13059, A
C 115	14.8	77.9	478	3	US-09-513-999C-11141	Sequence 11141, A	C 188	14.8	77.9	43307	3	US-08-949-016-15774	Sequence 15774, A
C 116	14.8	77.9	601	3	US-09-949-016-40853	Sequence 40853, A	C 189	14.8	77.9	64049	3	US-09-949-016-15744	Sequence 15744, A
C 117	14.8	77.9	601	3	US-09-949-016-50658	Sequence 50658, A	C 190	14.8	77.9	68667	3	US-09-949-016-17017	Sequence 17017, A
C 118	14.8	77.9	601	3	US-09-949-016-54367	Sequence 54367, A	C 191	14.8	77.9	70947	3	US-09-949-016-15165	Sequence 15165, A
C 119	14.8	77.9	601	3	US-09-949-016-54331	Sequence 54331, A	C 192	14.8	77.9	78157	3	US-09-949-016-16466	Sequence 16466, A
C 120	14.8	77.9	601	3	US-09-949-016-69064	Sequence 69064, A	C 193	14.8	77.9	78157	3	US-09-949-016-16467	Sequence 16467, A
C 121	14.8	77.9	601	3	US-09-949-016-70530	Sequence 70530, A	C 194	14.8	77.9	78649	3	US-08-949-016-14819	Sequence 14819, A
C 122	14.8	77.9	601	3	US-09-949-016-77508	Sequence 77508, A	C 195	14.8	77.9	78649	3	US-09-949-016-14820	Sequence 14820, A
C 123	14.8	77.9	601	3	US-09-949-016-77509	Sequence 77509, A	C 196	14.8	77.9	78649	3	US-09-949-016-16227	Sequence 16227, A
C 124	14.8	77.9	601	3	US-09-949-016-94629	Sequence 94629, A	C 197	14.8	77.9	78649	3	US-09-949-016-16328	Sequence 16328, A
C 125	14.8	77.9	601	3	US-09-949-016-122013	Sequence 122013, A	C 198	14.8	77.9	83218	3	US-09-949-016-14489	Sequence 14489, A
C 126	14.8	77.9	601	3	US-09-949-016-122014	Sequence 122014, A	C 199	14.8	77.9	83343	3	US-09-949-016-12242	Sequence 12242, A
C 127	14.8	77.9	601	3	US-09-949-016-134694	Sequence 134694, A	C 200	14.8	77.9	83497	3	US-09-949-016-15517	Sequence 15517, A
C 128	14.8	77.9	601	3	US-09-949-016-144401	Sequence 144401, A	C 201	14.8	77.9	85205	3	US-09-949-016-12211	Sequence 12211, A
C 129	14.8	77.9	601	3	US-09-949-016-167556	Sequence 167556, A	C 202	14.8	77.9	87205	3	US-09-949-016-13430	Sequence 13430, A
C 130	14.8	77.9	601	3	US-09-949-016-167663	Sequence 167663, A	C 203	14.8	77.9	90776	3	US-09-949-016-17230	Sequence 17230, A
C 131	14.8	77.9	601	3	US-09-949-016-181131	Sequence 181131, A	C 204	14.8	77.9	92407	3	US-09-596-002-36	Sequence 36, Appl
C 132	14.8	77.9	601	3	US-09-949-016-181132	Sequence 181132, A	C 205	14.8	77.9	93493	3	US-09-949-016-12063	Sequence 12063, A
C 133	14.8	77.9	601	3	US-09-949-016-181133	Sequence 181133, A	C 206	14.8	77.9	95020	3	US-09-949-016-13272	Sequence 13272, A
C 134	14.8	77.9	601	3	US-09-949-002-5993	Sequence 993, App	C 207	14.8	77.9	95318	3	US-09-949-016-13272	Sequence 13272, A
C 135	14.8	77.9	601	3	US-09-949-002-994	Sequence 994, App	C 208	14.8	77.9	95318	3	US-09-949-016-11784	Sequence 11784, A
C 136	14.8	77.9	601	3	US-09-949-002-4238	Sequence 4238, Ap	C 209	14.8	77.9	95561	3	US-09-949-016-12768	Sequence 12768, A
C 137	14.8	77.9	601	3	US-09-949-002-4239	Sequence 4239, Ap	C 210	14.8	77.9	95561	3	US-09-949-016-13306	Sequence 13306, A
C 138	14.8	77.9	662	3	US-09-270-767-28813	Sequence 28813, A	C 211	14.8	77.9	95561	3	US-09-949-016-13307	Sequence 13307, A
C 139	14.8	77.9	676	3	US-09-222-575-41	Sequence 41, Appl	C 212	14.8	77.9	96866	3	US-09-949-016-13768	Sequence 13768, A
C 140	14.8	77.9	676	3	US-09-389-681-41	Sequence 41, Appl	C 213	14.8	77.9	99916	3	US-09-816-095-3	Sequence 3, Appli
C 141	14.8	77.9	676	3	US-09-620-405B-41	Sequence 41, Appl	C 214	14.8	77.9	101786	4	US-09-531-120-199	Sequence 199, App
C 142	14.8	77.9	676	3	US-09-339-338-41	Sequence 41, Appl	C 215	14.8	77.9	105055	3	US-09-949-016-14001	Sequence 14001, A
C 143	14.8	77.9	676	3	US-09-433-826B-41	Sequence 41, Appl	C 216	14.8	77.9	109925	3	US-09-949-016-13210	Sequence 13210, A
C 144	14.8	77.9	676	3	US-09-604-287A-41	Sequence 41, Appl	C 217	14.8	77.9	11677	3	US-09-949-016-16946	Sequence 16946, A
C 145	14.8	77.9	676	3	US-09-285-480-41	Sequence 41, Appl	C 218	14.8	77.9	130563	3	US-09-949-016-12273	Sequence 12273, A
C 146	14.8	77.9	676	3	US-09-834-759-41	Sequence 41, Appl	C 219	14.8	77.9	130724	3	US-09-949-016-13753	Sequence 13753, A
C 147	14.8	77.9	676	3	US-09-530-751A-41	Sequence 41, Appl	C 220	14.8	77.9	131379	3	US-09-949-016-16050	Sequence 16050, A
C 148	14.8	77.9	676	3	US-09-551-621-41	Sequence 41, Appl	C 221	14.8	77.9	133559	3	US-09-949-016-15845	Sequence 15845, A
C 149	14.8	77.9	676	3	US-09-551-621A-41	Sequence 41, Appl	C 222	14.8	77.9	133613	3	US-09-949-016-15824	Sequence 15824, A
C 150	14.8	77.9	676	3	US-10-076-622-41	Sequence 41, Appl	C 223	14.8	77.9	141589	5	US-09-543-679A-2480	Sequence 2480, Ap
C 151	14.8	77.9	676	4	US-10-124-805-41	Sequence 41, Appl	C 224	14.8	77.9	141589	5	US-09-543-679A-2694	Sequence 2694, Ap
C 152	14.8	77.9	707	5	US-09-974-300-5294	Sequence 5294, Ap	C 225	14.8	77.9	141589	5	US-09-543-679A-2719	Sequence 2719, Ap
C 153	14.8	77.9	708	5	US-09-974-300-5134	Sequence 5134, Ap	C 226	14.8	77.9	146982	5	US-09-543-679A-3009	Sequence 3009, Ap
C 154	14.8	77.9	852	3	US-09-248-796A-2002	Sequence 2002, Ap	C 227	14.8	77.9	151828	4	US-09-531-120-197	Sequence 197, App
C 155	14.8	77.9	936	3	US-09-270-767-12940	Sequence 12940, A	C 228	14.8	77.9	161124	3	US-09-949-016-11760	Sequence 1760, A
C 156	14.8	77.9	966	3	US-08-748-506-8	Sequence 8, Appli	C 229	14.8	77.9	179304	3	US-09-128-155-17	Sequence 17, Appl
C 157	14.8	77.9	1001	3	US-09-641-638-388	Sequence 388, App	C 230	14.8	77.9	179904	3	US-09-949-002-577	Sequence 577, App
C 158	14.8	77.9	1001	3	US-09-641-638-390	Sequence 390, App	C 231	14.8	77.9	179905	3	US-09-949-002-705	Sequence 705, App
C 159	14.8	77.9	1001	3	US-10-170-097-388	Sequence 388, App	C 232	14.8	77.9	203475	3	US-09-949-016-14516	Sequence 14516, A
C 160	14.8	77.9	1001	3	US-10-170-097-390	Sequence 390, App	C 233	14.8	77.9	203475	3	US-09-949-016-14517	Sequence 14517, A
C 161	14.8	77.9	1177	2	US-08-249-671A-8	Sequence 8, Appli	C 234	14.8	77.9	203475	3	US-09-949-016-14518	Sequence 14518, A
C 162	14.8	77.9	1236	3	US-09-799-451-867	Sequence 867, App	C 235	14.8	77.9	203475	3	US-09-949-016-14519	Sequence 14519, A
C 163	14.8	77.9	1320	3	US-09-023-655-491	Sequence 491, App	C 236	14.8	77.9	203475	3	US-09-949-016-17226	Sequence 17226, A
C 164	14.8	77.9	2102	3	US-09-023-655-364	Sequence 364, App	C 237	14.8	77.9	203475	3	US-09-949-016-17227	Sequence 17227, A
C 165	14.8	77.9	2313	3	US-09-107-532A-734	Sequence 734, App	C 238	14.8	77.9	203475	3	US-09-949-016-17228	Sequence 17228, A
C 166	14.8	77.9	2457	4	US-10-094-749-438	Sequence 438, App	C 239	14.8	77.9	203475	3	US-09-949-016-17229	Sequence 17229, A
C 167	14.8	77.9	2760	3	US-09-543-681A-86	Sequence 86, Appl	C 240	14.8	77.9	209274	5	US-09-543-679A-3004	Sequence 3004, Ap
C 168	14.8	77.9	3089	3	US-09-016-434-1330	Sequence 1330, Ap	C 241	14.8	77.9	260286	3	US-09-949-016-17037	Sequence 17037, A
C 169	14.8	77.9	3089	4	US-09-880-107-3687	Sequence 3687, Ap	C 242	14.8	77.9	260293	3	US-09-949-016-12106	Sequence 12106, A



243	14.8	77.9	267482	3	US-09-949-002-659	Sequence 659, App	C 316	14.2	74.7	131	3	US-10-150-104-1	Sequence 1, Appli
244	14.8	77.9	267505	3	US-09-949-002-783	Sequence 783, App	317	14.2	74.7	267	5	US-09-974-300-6246	Sequence 6246, Ap
245	14.8	77.9	392000	3	US-10-027-983-11	Sequence 11, Appl	C 318	14.2	74.7	198	3	US-09-313-294A-1237	Sequence 1237, Ap
246	14.8	77.9	462589	3	US-09-949-016-12900	Sequence 12900, A	C 319	14.2	74.7	300	4	US-09-297-648-2319	Sequence 2319, Ap
247	14.8	77.9	476044	3	US-09-949-016-12412	Sequence 12412, A	C 320	14.2	74.7	318	3	US-09-221-017B-458	Sequence 458, App
248	14.8	77.9	611587	4	US-09-531-120-209	Sequence 209, App	C 321	14.2	74.7	360	3	US-09-270-767-29190	Sequence 29190, A
249	14.8	77.9	678533	3	US-09-949-016-14577	Sequence 14577, A	C 322	14.2	74.7	378	3	US-10-002-623-563	Sequence 563, App
250	14.8	77.9	678533	3	US-09-949-016-14578	Sequence 14578, A	C 323	14.2	74.7	449	3	US-09-513-999C-917	Sequence 917, App
251	14.8	77.9	1664976	3	US-08-916-421B-1	Sequence 1, Appli	C 324	14.2	74.7	476	3	US-09-513-999C-11720	Sequence 11720, A
252	14.8	77.9	1664976	3	US-09-692-570-1	Sequence 1, Appli	C 325	14.2	74.7	480	3	US-09-198-930-206	Sequence 206, App
C 253	14.8	77.9	1830121	3	US-09-557-884-1	Sequence 1, Appli	326	14.2	74.7	480	3	US-09-312-283C-206	Sequence 206, App
C 254	14.8	77.9	1830121	3	US-09-643-990A-1	Sequence 1, Appli	327	14.2	74.7	480	3	US-09-107-532A-420	Sequence 420, App
C 255	14.8	77.9	1830121	3	US-10-158-865-1	Sequence 1, Appli	C 328	14.2	74.7	494	3	US-09-513-999C-892	Sequence 892, App
256	14.4	75.8	47	3	US-09-641-638-1043	Sequence 1043, Ap	C 329	14.2	74.7	549	3	US-09-854-133-515	Sequence 515, App
257	14.4	75.8	312	3	US-10-170-097-1043	Sequence 1043, Ap	C 330	14.2	74.7	573	3	US-09-248-796A-3436	Sequence 3436, Ap
C 258	14.4	75.8	312	3	US-09-248-796A-2555	Sequence 2555, Ap	C 331	14.2	74.7	601	3	US-09-949-016-58296	Sequence 58296, A
C 259	14.4	75.8	356	2	US-08-606-789-15	Sequence 15, Appl	C 332	14.2	74.7	601	3	US-09-949-016-60332	Sequence 60332, A
260	14.4	75.8	356	2	US-09-111-348-15	Sequence 15, Appl	C 333	14.2	74.7	601	3	US-09-949-016-74393	Sequence 74393, A
261	14.4	75.8	437	4	US-09-880-107-2647	Sequence 2647, Ap	C 334	14.2	74.7	601	3	US-09-949-016-76486	Sequence 76486, A
C 262	14.4	75.8	565	3	US-09-257-584-10	Sequence 10, Appl	C 335	14.2	74.7	601	3	US-09-949-016-89613	Sequence 89613, A
C 263	14.4	75.8	601	3	US-09-949-016-44349	Sequence 44349, A	C 336	14.2	74.7	601	3	US-09-949-016-89614	Sequence 89614, A
C 264	14.4	75.8	601	3	US-09-949-016-57961	Sequence 57961, A	C 337	14.2	74.7	601	3	US-09-949-016-112118	Sequence 112118, A
C 265	14.4	75.8	601	3	US-09-949-016-57962	Sequence 57962, A	C 338	14.2	74.7	601	3	US-09-949-016-119242	Sequence 119242, A
266	14.4	75.8	601	3	US-09-949-016-64318	Sequence 64318, A	C 339	14.2	74.7	601	3	US-09-949-016-119373	Sequence 119373, A
C 267	14.4	75.8	601	3	US-09-949-016-76381	Sequence 76381, A	C 340	14.2	74.7	601	3	US-09-949-016-123966	Sequence 123966, A
C 268	14.4	75.8	601	3	US-09-949-016-76382	Sequence 76382, A	C 341	14.2	74.7	601	3	US-09-949-016-126758	Sequence 126758, A
269	14.4	75.8	601	3	US-09-949-016-87508	Sequence 87508, A	C 342	14.2	74.7	601	3	US-09-949-016-126759	Sequence 126759, A
C 270	14.4	75.8	601	3	US-09-949-016-124311	Sequence 124311, A	C 343	14.2	74.7	601	3	US-09-949-016-126807	Sequence 126807, A
C 271	14.4	75.8	601	3	US-09-949-016-144426	Sequence 144426, A	C 344	14.2	74.7	601	3	US-09-949-016-126808	Sequence 126808, A
C 272	14.4	75.8	601	3	US-09-949-016-150280	Sequence 150280, A	C 345	14.2	74.7	601	3	US-09-949-016-126856	Sequence 126856, A
C 273	14.4	75.8	601	3	US-09-949-016-177324	Sequence 177324, A	C 346	14.2	74.7	601	3	US-09-949-016-126857	Sequence 126857, A
C 274	14.4	75.8	801	5	US-09-974-300-8142	Sequence 8142, Ap	C 347	14.2	74.7	601	3	US-09-949-016-133593	Sequence 133593, A
275	14.4	75.8	1001	3	US-09-641-638-389	Sequence 389, App	C 348	14.2	74.7	601	3	US-09-949-016-134391	Sequence 134391, A
276	14.4	75.8	1001	3	US-10-170-097-389	Sequence 389, App	C 349	14.2	74.7	601	3	US-09-949-016-134392	Sequence 134392, A
C 277	14.4	75.8	1082	3	US-09-484-970B-109	Sequence 109, App	C 350	14.2	74.7	601	3	US-09-949-016-134440	Sequence 134440, A
C 278	14.4	75.8	1185	3	US-09-248-796A-1925	Sequence 1925, Ap	C 351	14.2	74.7	601	3	US-09-949-016-134441	Sequence 134441, A
C 279	14.4	75.8	1563	3	US-09-377-557-11	Sequence 11, Appl	C 352	14.2	74.7	601	3	US-09-949-016-134489	Sequence 134489, A
C 280	14.4	75.8	2706	2	US-08-630-822A-61	Sequence 61, Appl	C 353	14.2	74.7	601	3	US-09-949-016-134490	Sequence 134490, A
C 281	14.4	75.8	2706	2	US-09-005-069-61	Sequence 61, Appl	C 354	14.2	74.7	601	3	US-09-949-016-152082	Sequence 152082, A
C 282	14.4	75.8	2706	3	US-09-171-156A-20	Sequence 20, Appl	C 355	14.2	74.7	601	3	US-09-949-016-152083	Sequence 152083, A
C 283	14.4	75.8	2706	3	US-09-004-730A-20	Sequence 20, Appl	C 356	14.2	74.7	601	3	US-09-949-016-152084	Sequence 152084, A
C 284	14.4	75.8	2706	3	US-08-981-799A-20	Sequence 20, Appl	C 357	14.2	74.7	601	3	US-09-949-016-152155	Sequence 152155, A
C 285	14.4	75.8	2847	3	US-09-789-451-513	Sequence 513, App	C 358	14.2	74.7	601	3	US-09-949-016-152156	Sequence 152156, A
C 286	14.4	75.8	8640	3	US-09-964-956-24	Sequence 24, Appl	C 359	14.2	74.7	601	3	US-09-949-016-152157	Sequence 152157, A
287	14.4	75.8	8640	3	US-09-964-956-26	Sequence 26, Appl	C 360	14.2	74.7	601	3	US-09-949-016-152228	Sequence 152228, A
C 288	14.4	75.8	18367	3	US-09-949-016-14065	Sequence 14065, A	C 361	14.2	74.7	601	3	US-09-949-016-152229	Sequence 152229, A
C 289	14.4	75.8	21516	3	US-09-949-016-13024	Sequence 13024, A	C 362	14.2	74.7	601	3	US-09-949-016-152230	Sequence 152230, A
290	14.4	75.8	29636	3	US-09-949-016-12189	Sequence 12189, A	C 363	14.2	74.7	601	3	US-09-949-016-152301	Sequence 152301, A
291	14.4	75.8	29637	3	US-09-949-016-17173	Sequence 17173, A	C 364	14.2	74.7	601	3	US-09-949-016-152302	Sequence 152302, A
292	14.4	75.8	29638	3	US-09-949-016-14137	Sequence 14137, A	C 365	14.2	74.7	601	3	US-09-949-016-152303	Sequence 152303, A
293	14.4	75.8	33478	3	US-09-949-016-13615	Sequence 13615, A	C 366	14.2	74.7	601	3	US-09-949-016-154785	Sequence 154785, A
C 294	14.4	75.8	36047	3	US-09-949-016-13966	Sequence 13966, A	C 367	14.2	74.7	601	3	US-09-949-016-159005	Sequence 159005, A
C 295	14.4	75.8	47968	3	US-09-949-016-15240	Sequence 15240, A	C 368	14.2	74.7	601	3	US-09-949-016-159387	Sequence 159387, A
296	14.4	75.8	68173	3	US-09-949-016-14046	Sequence 14046, A	C 369	14.2	74.7	601	3	US-09-949-016-159388	Sequence 159388, A
297	14.4	75.8	69909	3	US-09-949-016-13423	Sequence 13423, A	C 370	14.2	74.7	601	3	US-09-949-016-159389	Sequence 159389, A
298	14.4	75.8	77535	3	US-09-949-016-14279	Sequence 14279, A	C 371	14.2	74.7	601	3	US-09-949-016-159460	Sequence 159460, A
299	14.4	75.8	77535	3	US-09-949-016-14280	Sequence 14280, A	C 372	14.2	74.7	601	3	US-09-949-016-159461	Sequence 159461, A
300	14.4	75.8	77535	3	US-09-949-016-14281	Sequence 14281, A	C 373	14.2	74.7	601	3	US-09-949-016-159462	Sequence 159462, A
C 301	14.4	75.8	84252	3	US-09-949-016-17315	Sequence 17315, A	C 374	14.2	74.7	601	3	US-09-949-016-159533	Sequence 159533, A
C 302	14.4	75.8	96048	3	US-09-949-016-15966	Sequence 15966, A	C 375	14.2	74.7	601	3	US-09-949-016-159534	Sequence 159534, A
C 303	14.4	75.8	133613	3	US-09-949-016-15824	Sequence 15824, A	C 376	14.2	74.7	601	3	US-09-949-016-159535	Sequence 159535, A
C 304	14.4	75.8	144158	3	US-09-949-016-11755	Sequence 11755, A	C 377	14.2	74.7	601	3	US-09-949-016-159606	Sequence 159606, A
C 305	14.4	75.8	144158	3	US-09-949-016-12936	Sequence 12936, A	C 378	14.2	74.7	601	3	US-09-949-016-159607	Sequence 159607, A
306	14.4	75.8	145241	3	US-09-949-016-17394	Sequence 17394, A	C 379	14.2	74.7	601	3	US-09-949-016-159608	Sequence 159608, A
307	14.4	75.8	145241	3	US-09-949-016-17395	Sequence 17395, A	C 380	14.2	74.7	601	3	US-09-949-016-161881	Sequence 161881, A
308	14.4	75.8	176006	3	US-09-949-016-16804	Sequence 16804, A	C 381	14.2	74.7	601	3	US-09-949-016-166114	Sequence 166114, A
309	14.4	75.8	187580	3	US-09-949-016-13266	Sequence 13266, A	C 382	14.2	74.7	601	3	US-09-949-016-171581	Sequence 171581, A
310	14.4	75.8	192700	3	US-09-949-016-11820	Sequence 11820, A	C 383	14.2	74.7	601	3	US-09-949-016-179311	Sequence 179311, A
311	14.4	75.8	192704	3	US-09-949-016-17182	Sequence 17182, A	C 384	14.2	74.7	601	3	US-09-949-016-179312	Sequence 179312, A
312	14.4	75.8	233375	3	US-09-949-016-12849	Sequence 12849, A	C 385	14.2	74.7	601	3	US-09-949-016-186673	Sequence 186673, A
313	14.4	75.8	283538	3	US-09-949-016-13506	Sequence 13506, A	C 386	14.2	74.7	601	3	US-09-949-016-186674	Sequence 186674, A
314	14.4	75.8	670689	3	US-09-949-016-12505	Sequence 12505, A	C 387	14.2	74.7	601	3	US-09-949-016-186675	Sequence 186675, A
315	14.4	75.8	670690	3	US-09-949-016-14207	Sequence 14207, A	C 388	14.2	74.7	601	3	US-09-949-016-201189	Sequence 201189, A

389	14.2	74.7	601	3	US-09-949-016-205561	Sequence 205561,	C 462	14.2	74.7	4680	2	US-08-254-358-1	Sequence 1, Appli
C 390	14.2	74.7	601	3	US-09-949-016-206956	Sequence 206956,	C 463	14.2	74.7	4680	2	US-08-475-391-1	Sequence 1, Appli
C 391	14.2	74.7	601	3	US-09-949-016-206957	Sequence 206957,	C 464	14.2	74.7	4680	2	US-08-709-609-1	Sequence 1, Appli
C 392	14.2	74.7	601	3	US-09-949-016-206958	Sequence 206958,	C 465	14.2	74.7	4680	7	PCT-US95-07178-1	Sequence 1, Appli
C 393	14.2	74.7	601	3	US-09-949-002-13904	Sequence 13904, Ap	C 466	14.2	74.7	4681	3	US-09-807-802A-18	Sequence 18, Appl
C 394	14.2	74.7	601	3	US-09-949-002-13905	Sequence 13905, Ap	C 467	14.2	74.7	4800	3	US-09-949-016-15055	Sequence 15055, A
C 395	14.2	74.7	601	3	US-09-949-002-5851	Sequence 5851, Ap	C 468	14.2	74.7	4833	3	US-09-949-016-14282	Sequence 14282, A
C 396	14.2	74.7	601	3	US-09-949-002-5852	Sequence 5852, Ap	C 469	14.2	74.7	5761	2	US-07-749-001-2	Sequence 2, Appli
C 397	14.2	74.7	611	3	US-09-385-982-378	Sequence 378, App	C 470	14.2	74.7	5761	2	US-08-154-198-2	Sequence 2, Appli
C 398	14.2	74.7	603	5	US-09-974-300-507	Sequence 507, App	C 471	14.2	74.7	5761	2	US-08-463-335-2	Sequence 2, Appli
C 399	14.2	74.7	700	3	US-09-735-271-266	Sequence 266, App	C 472	14.2	74.7	5761	2	US-08-464-023A-2	Sequence 2, Appli
C 400	14.2	74.7	708	3	US-08-533-559-7504	Sequence 7504, Ap	C 473	14.2	74.7	5922	3	US-08-961-527-267	Sequence 267, App
C 401	14.2	74.7	715	3	US-08-998-416-922	Sequence 922, App	C 474	14.2	74.7	6708	3	US-09-949-016-12933	Sequence 12933, A
C 402	14.2	74.7	728	3	US-09-988-863A-4	Sequence 4, Appli	C 475	14.2	74.7	7557	3	US-09-770-315-3	Sequence 3, Appli
C 403	14.2	74.7	749	3	US-09-270-767-13258	Sequence 13258, A	C 476	14.2	74.7	8151	3	US-09-438-268-2	Sequence 2, Appli
C 404	14.2	74.7	778	4	US-09-297-648-3659	Sequence 3659, Ap	C 477	14.2	74.7	8201	4	US-09-958-617A-17	Sequence 17, Appl
C 405	14.2	74.7	800	3	US-09-949-016-3313	Sequence 3313, Ap	C 478	14.2	74.7	8698	3	US-09-770-315-2	Sequence 2, Appli
C 406	14.2	74.7	811	3	US-09-949-016-3371	Sequence 3371, Ap	C 479	14.2	74.7	8746	3	US-09-949-016-16984	Sequence 16984, A
C 407	14.2	74.7	820	3	US-09-949-016-2540	Sequence 2540, Ap	C 480	14.2	74.7	9192	3	US-09-949-016-14830	Sequence 14830, A
C 408	14.2	74.7	896	3	US-09-188-930-36	Sequence 36, Appl	C 481	14.2	74.7	10207	2	US-08-920-812-2	Sequence 2, Appli
C 409	14.2	74.7	896	3	US-09-312-283C-36	Sequence 36, Appl	C 482	14.2	74.7	10207	2	US-08-920-827-2	Sequence 2, Appli
C 410	14.2	74.7	903	5	US-09-974-300-1703	Sequence 1703, Ap	C 483	14.2	74.7	10207	2	US-08-921-177-2	Sequence 2, Appli
C 411	14.2	74.7	975	3	US-09-489-039A-1422	Sequence 1422, Ap	C 484	14.2	74.7	10207	2	US-08-362-577C-2	Sequence 2, Appli
C 412	14.2	74.7	983	3	US-08-270-767-11267	Sequence 11267, A	C 485	14.2	74.7	10207	2	US-08-920-828-2	Sequence 2, Appli
C 413	14.2	74.7	990	3	US-09-134-000C-1367	Sequence 1367, Ap	C 486	14.2	74.7	11728	3	US-09-949-016-14245	Sequence 14245, A
C 414	14.2	74.7	1049	3	US-09-799-451-892	Sequence 892, App	C 487	14.2	74.7	12139	3	US-09-949-016-14218	Sequence 14218, A
C 415	14.2	74.7	1107	5	US-09-974-300-1582	Sequence 1582, Ap	C 488	14.2	74.7	12193	3	US-09-949-016-16237	Sequence 16237, A
C 416	14.2	74.7	1215	3	US-09-328-352-2728	Sequence 2728, Ap	C 489	14.2	74.7	12423	3	US-09-949-016-14959	Sequence 14959, A
C 417	14.2	74.7	1317	3	US-09-350-756-5	Sequence 5, Appli	C 490	14.2	74.7	12616	3	US-09-949-016-14217	Sequence 14217, A
C 418	14.2	74.7	1323	3	US-09-583-110-1395	Sequence 1395, Ap	C 491	14.2	74.7	16995	3	US-08-961-527-82	Sequence 82, Appl
C 419	14.2	74.7	1338	3	US-09-107-433-915	Sequence 915, App	C 492	14.2	74.7	18026	3	US-09-949-016-13309	Sequence 13309, A
C 420	14.2	74.7	1352	3	US-09-721-870-15	Sequence 15, Appl	C 493	14.2	74.7	19062	5	US-09-543-679A-2847	Sequence 2847, Ap
C 421	14.2	74.7	1386	3	US-09-134-000C-2146	Sequence 2146, Ap	C 494	14.2	74.7	19389	3	US-09-949-016-15113	Sequence 15113, A
C 422	14.2	74.7	1446	3	US-09-252-991A-16240	Sequence 16240, A	C 495	14.2	74.7	20079	3	US-09-949-016-16060	Sequence 16060, A
C 423	14.2	74.7	1500	3	US-09-252-991A-16351	Sequence 16351, A	C 496	14.2	74.7	25992	3	US-09-949-016-17308	Sequence 17308, A
C 424	14.2	74.7	1842	3	US-09-016-434-46	Sequence 46, Appl	C 497	14.2	74.7	26000	3	US-09-843-376-10	Sequence 10, Appl
C 425	14.2	74.7	1887	4	US-10-099-322-1	Sequence 1, Appli	C 498	14.2	74.7	30417	5	US-09-543-679A-2849	Sequence 2849, Ap
C 426	14.2	74.7	1887	4	US-10-094-564-1	Sequence 1, Appli	C 499	14.2	74.7	31618	3	US-09-949-016-14759	Sequence 14759, Ap
C 427	14.2	74.7	1973	10	5256643-1	Patent No. 5256643	C 500	14.2	74.7	36156	3	US-09-949-016-12128	Sequence 12128, A
C 428	14.2	74.7	2033	3	US-09-023-655-1266	Sequence 1266, Ap	C 501	14.2	74.7	36156	3	US-09-949-016-13261	Sequence 13261, A
C 429	14.2	74.7	2033	3	US-09-949-016-5242	Sequence 5242, Ap	C 502	14.2	74.7	37925	5	US-09-984-429-701	Sequence 701, App
C 430	14.2	74.7	2160	3	US-09-107-532A-2338	Sequence 2338, Ap	C 503	14.2	74.7	44166	3	US-09-949-016-15829	Sequence 15829, A
C 431	14.2	74.7	2184	4	US-10-094-749-322	Sequence 322, App	C 504	14.2	74.7	45432	3	US-09-949-002-719	Sequence 719, App
C 432	14.2	74.7	2251	3	US-10-104-047-916	Sequence 916, App	C 505	14.2	74.7	47084	3	US-09-949-016-13380	Sequence 13380, A
C 433	14.2	74.7	2361	3	US-09-107-433-1290	Sequence 1290, Ap	C 506	14.2	74.7	47110	3	US-09-949-016-17609	Sequence 17609, A
C 434	14.2	74.7	2370	4	US-10-094-749-2	Sequence 2, Appli	C 507	14.2	74.7	51719	3	US-09-918-686-2	Sequence 2, Appli
C 435	14.2	74.7	2427	3	US-09-543-681A-823	Sequence 823, App	C 508	14.2	74.7	54180	3	US-09-949-016-14894	Sequence 14894, A
C 436	14.2	74.7	2479	3	US-09-949-016-3088	Sequence 3088, Ap	C 509	14.2	74.7	61158	3	US-09-949-016-15041	Sequence 15041, A
C 437	14.2	74.7	2519	5	US-10-770-127-202	Sequence 202, App	C 510	14.2	74.7	62311	3	US-09-949-016-14582	Sequence 14582, A
C 438	14.2	74.7	2520	3	US-09-897-427A-3	Sequence 3, Appli	C 511	14.2	74.7	70323	3	US-09-949-016-17594	Sequence 17594, A
C 439	14.2	74.7	2581	3	US-09-363-708-1	Sequence 1, Appli	C 512	14.2	74.7	75431	3	US-09-949-016-15122	Sequence 15122, A
C 440	14.2	74.7	2581	3	US-09-083-587-1	Sequence 1, Appli	C 513	14.2	74.7	76553	3	US-09-949-016-13432	Sequence 13432, A
C 441	14.2	74.7	2604	3	US-09-248-796A-5104	Sequence 5104, Ap	C 514	14.2	74.7	78846	3	US-09-949-016-12396	Sequence 12396, A
C 442	14.2	74.7	2667	2	US-09-614-221A-335	Sequence 335, App	C 515	14.2	74.7	78846	3	US-09-949-016-12791	Sequence 12791, A
C 443	14.2	74.7	2675	2	US-07-749-001-4	Sequence 4, Appli	C 516	14.2	74.7	78846	3	US-09-949-016-12792	Sequence 12792, A
C 444	14.2	74.7	2675	2	US-08-154-198-4	Sequence 4, Appli	C 517	14.2	74.7	78846	3	US-09-949-016-12793	Sequence 12793, A
C 445	14.2	74.7	2675	2	US-08-463-335-4	Sequence 4, Appli	C 518	14.2	74.7	78850	3	US-09-949-016-16013	Sequence 16013, A
C 446	14.2	74.7	2675	2	US-08-464-023A-4	Sequence 4, Appli	C 519	14.2	74.7	78850	3	US-09-949-016-16014	Sequence 16014, A
C 447	14.2	74.7	2708	3	US-09-949-016-1191	Sequence 1191, Ap	C 520	14.2	74.7	78850	3	US-09-949-016-16015	Sequence 16015, A
C 448	14.2	74.7	2718	3	US-09-248-796A-2230	Sequence 2230, Ap	C 521	14.2	74.7	78850	3	US-09-949-016-16016	Sequence 16016, A
C 449	14.2	74.7	2767	3	US-09-306-595C-4	Sequence 4, Appli	C 522	14.2	74.7	78850	3	US-09-949-016-16201	Sequence 16201, A
C 450	14.2	74.7	2767	3	US-09-925-388-4	Sequence 4, Appli	C 523	14.2	74.7	78850	3	US-09-949-016-16202	Sequence 16202, A
C 451	14.2	74.7	2866	3	US-10-431-846-4	Sequence 4, Appli	C 524	14.2	74.7	78850	3	US-09-949-016-16203	Sequence 16203, A
C 452	14.2	74.7	2893	3	US-10-104-047-1267	Sequence 1267, Ap	C 525	14.2	74.7	78850	3	US-09-949-016-16204	Sequence 16204, A
C 453	14.2	74.7	2998	4	US-10-094-749-651	Sequence 651, App	C 526	14.2	74.7	79578	3	US-09-949-016-16339	Sequence 16339, A
C 454	14.2	74.7	3138	3	US-10-104-047-1676	Sequence 1676, Ap	C 527	14.2	74.7	79595	3	US-09-949-016-15318	Sequence 15318, A
C 455	14.2	74.7	4072	3	US-09-770-315-4	Sequence 4, Appli	C 528	14.2	74.7	80490	3	US-09-949-002-733	Sequence 733, App
C 456	14.2	74.7	4407	3	US-09-949-016-1690	Sequence 1690, Ap	C 529	14.2	74.7	82048	3	US-09-949-002-696	Sequence 696, App
C 457	14.2	74.7	4675	3	US-09-782-378A-1	Sequence 1, Appli	C 530	14.2	74.7	82619	3	US-09-949-002-579	Sequence 579, App
C 458	14.2	74.7	4675	3	US-09-782-378A-2	Sequence 2, Appli	C 531	14.2	74.7	85122	3	US-09-949-016-14693	Sequence 14693, A
C 459	14.2	74.7	4675	4	US-10-111-708-1	Sequence 1, Appli	C 532	14.2	74.7	87594	3	US-09-949-016-12135	Sequence 12135, A
C 460	14.2	74.7	4679	3	US-10-038-972A-12	Sequence 12, Appl	C 533	14.2	74.7	87611	3	US-09-949-016-16139	Sequence 16139, A
C 461	14.2	74.7	4679	4	US-09-717-789C-25	Sequence 25, Appl	C 534	14.2	74.7	92119	3	US-09-918-686-1	Sequence 1, Appli

C 535	14.2	74.7	94142	3	US-09-949-016-16553	Sequence 16553, A	608	13.8	72.6	330	3	US-10-017-754-1379	Sequence 1379, Ap
C 536	14.2	74.7	94618	4	US-09-531-120-191	Sequence 191, App	609	13.8	72.6	330	3	US-09-651-563-1379	Sequence 1379, Ap
C 537	14.2	74.7	96922	4	US-09-949-016-17061	Sequence 17061, A	C 610	13.8	72.6	342	3	US-09-134-000C-2932	Sequence 2932, Ap
C 538	14.2	74.7	10786	4	US-09-531-120-199	Sequence 199, App	C 611	13.8	72.6	346	2	US-08-967-101-33	Sequence 33, Appl
C 539	14.2	74.7	105919	4	US-09-949-016-11769	Sequence 11769, A	C 612	13.8	72.6	346	2	US-08-592-541-33	Sequence 33, Appl
C 540	14.2	74.7	110096	4	US-09-880-107-1542	Sequence 1542, Ap	C 613	13.8	72.6	346	2	US-09-124-698-33	Sequence 33, Appl
C 541	14.2	74.7	116652	4	US-09-949-016-13413	Sequence 13413, A	C 614	13.8	72.6	346	3	US-09-127-480-33	Sequence 33, Appl
C 542	14.2	74.7	119214	3	US-09-949-016-12507	Sequence 12507, A	C 615	13.8	72.6	346	3	US-08-496-841C-33	Sequence 33, Appl
C 543	14.2	74.7	120609	3	US-09-949-016-13915	Sequence 13915, A	C 616	13.8	72.6	346	3	US-09-124-523-33	Sequence 33, Appl
C 544	14.2	74.7	128516	3	US-09-949-016-13501	Sequence 13501, A	C 617	13.8	72.6	346	3	US-09-636-796A-33	Sequence 33, Appl
C 545	14.2	74.7	129778	3	US-09-949-016-12191	Sequence 12191, A	C 618	13.8	72.6	346	3	US-08-431-048F-33	Sequence 33, Appl
C 546	14.2	74.7	129778	3	US-09-949-016-17075	Sequence 17075, A	C 619	13.8	72.6	346	4	US-09-689-159A-33	Sequence 33, Appl
C 547	14.2	74.7	133157	3	US-09-949-016-12541	Sequence 12541, A	C 620	13.8	72.6	353	3	US-09-513-999C-13247	Sequence 13247, A
C 548	14.2	74.7	134987	3	US-09-949-016-15348	Sequence 15348, A	C 621	13.8	72.6	381	5	US-09-974-300-1435	Sequence 1435, Ap
C 549	14.2	74.7	134987	3	US-09-949-016-15349	Sequence 15349, A	C 622	13.8	72.6	399	3	US-09-854-133-701	Sequence 701, App
C 550	14.2	74.7	134987	3	US-09-949-016-15350	Sequence 15350, A	C 623	13.8	72.6	429	3	US-09-107-532A-1963	Sequence 1963, Ap
C 551	14.2	74.7	134987	3	US-09-949-016-15507	Sequence 15507, A	C 624	13.8	72.6	459	3	US-09-621-976-16276	Sequence 16276, A
C 552	14.2	74.7	134987	3	US-09-949-016-15508	Sequence 15508, A	C 625	13.8	72.6	462	3	US-09-401-064-164	Sequence 164, App
C 553	14.2	74.7	135667	3	US-09-949-016-15509	Sequence 15509, A	C 626	13.8	72.6	468	3	US-09-513-999C-12568	Sequence 12568, A
C 554	14.2	74.7	135667	3	US-09-949-016-15501	Sequence 15051, A	C 627	13.8	72.6	471	2	US-08-502-535B-1	Sequence 1, Appli
C 555	14.2	74.7	139552	3	US-09-949-016-15300	Sequence 15300, A	C 628	13.8	72.6	471	2	US-08-908-005A-1	Sequence 1, Appli
C 556	14.2	74.7	147840	3	US-09-949-016-15236	Sequence 15236, A	C 629	13.8	72.6	471	3	US-08-588-818-6	Sequence 6, Appli
C 557	14.2	74.7	147894	3	US-09-949-016-12751	Sequence 12751, A	C 630	13.8	72.6	471	3	US-08-974-469A-6	Sequence 6, Appli
C 558	14.2	74.7	150023	3	US-09-949-016-17057	Sequence 17057, A	C 631	13.8	72.6	471	3	US-08-832-180-3	Sequence 3, Appli
C 559	14.2	74.7	155617	3	US-09-949-016-16191	Sequence 16191, A	C 632	13.8	72.6	471	3	US-08-832-198-10	Sequence 10, Appl
C 560	14.2	74.7	163664	3	US-09-949-016-12545	Sequence 12545, A	C 633	13.8	72.6	471	3	US-09-253-523-1	Sequence 1, Appli
C 561	14.2	74.7	163664	3	US-09-949-016-12546	Sequence 12546, A	C 634	13.8	72.6	471	3	US-09-251-911-1	Sequence 1, Appli
C 562	14.2	74.7	174029	3	US-09-949-016-12610	Sequence 12610, A	C 635	13.8	72.6	471	3	US-09-819-902-10	Sequence 10, Appl
C 563	14.2	74.7	174030	3	US-09-949-016-13880	Sequence 13880, A	C 636	13.8	72.6	471	3	US-09-752-510-10	Sequence 10, Appl
C 564	14.2	74.7	189560	3	US-09-949-016-17202	Sequence 17202, A	C 637	13.8	72.6	471	3	US-08-982-285-22	Sequence 22, Appl
C 565	14.2	74.7	198632	3	US-09-949-016-12781	Sequence 12781, A	C 638	13.8	72.6	471	3	US-08-982-285-23	Sequence 23, Appl
C 566	14.2	74.7	198637	3	US-09-949-016-17393	Sequence 17393, A	C 639	13.8	72.6	471	3	US-08-982-285-28	Sequence 28, Appl
C 567	14.2	74.7	209631	3	US-09-949-002-574	Sequence 574, App	C 640	13.8	72.6	471	3	US-08-711-899-6	Sequence 6, Appli
C 568	14.2	74.7	209632	3	US-09-949-002-802	Sequence 802, App	C 641	13.8	72.6	471	3	US-10-100-057-18	Sequence 18, Appl
C 569	14.2	74.7	219964	3	US-09-949-016-15086	Sequence 15086, A	C 642	13.8	72.6	474	3	US-10-105-080-2	Sequence 2, Appli
C 570	14.2	74.7	234884	3	US-09-949-016-16420	Sequence 16420, A	C 643	13.8	72.6	474	3	US-10-105-080-9	Sequence 9, Appli
C 571	14.2	74.7	237510	3	US-09-949-016-14273	Sequence 14273, A	C 644	13.8	72.6	513	3	US-09-710-279-1925	Sequence 1925, Ap
C 572	14.2	74.7	238815	3	US-09-949-016-16274	Sequence 16274, A	C 645	13.8	72.6	541	3	US-08-679-409-39	Sequence 39, Appl
C 573	14.2	74.7	245286	3	US-09-949-016-15497	Sequence 15497, A	C 646	13.8	72.6	570	3	US-08-982-285-29	Sequence 29, Appl
C 574	14.2	74.7	256171	3	US-09-949-016-12822	Sequence 12822, A	C 647	13.8	72.6	579	3	US-10-105-080-7	Sequence 7, Appl
C 575	14.2	74.7	256175	3	US-09-949-016-15524	Sequence 15524, A	C 648	13.8	72.6	601	3	US-09-949-016-20143	Sequence 20143, A
C 576	14.2	74.7	256776	3	US-09-949-016-16435	Sequence 16435, A	C 649	13.8	72.6	601	3	US-09-949-016-23899	Sequence 23899, A
C 577	14.2	74.7	263693	3	US-09-949-016-12386	Sequence 12386, A	C 650	13.8	72.6	601	3	US-09-949-016-23900	Sequence 23900, A
C 578	14.2	74.7	263694	3	US-09-949-016-16915	Sequence 16915, A	C 651	13.8	72.6	601	3	US-09-949-016-23901	Sequence 23901, A
C 579	14.2	74.7	271134	3	US-09-949-016-12705	Sequence 12705, A	C 652	13.8	72.6	601	3	US-09-949-016-25725	Sequence 25725, A
C 580	14.2	74.7	275110	3	US-09-949-016-12706	Sequence 12706, A	C 653	13.8	72.6	601	3	US-09-949-016-31201	Sequence 31201, A
C 581	14.2	74.7	275110	3	US-09-949-016-16700	Sequence 16700, A	C 654	13.8	72.6	601	3	US-09-949-016-31457	Sequence 31457, A
C 582	14.2	74.7	298336	3	US-09-949-016-16600	Sequence 16600, A	C 655	13.8	72.6	601	3	US-09-949-016-31458	Sequence 31458, A
C 583	14.2	74.7	301828	3	US-09-949-016-13969	Sequence 13969, A	C 656	13.8	72.6	601	3	US-09-949-016-31459	Sequence 31459, A
C 584	14.2	74.7	305491	3	US-09-949-016-17550	Sequence 17550, A	C 657	13.8	72.6	601	3	US-09-949-016-33735	Sequence 33735, A
C 585	14.2	74.7	317366	3	US-09-949-016-16001	Sequence 16001, A	C 658	13.8	72.6	601	3	US-09-949-016-34165	Sequence 34165, A
C 586	14.2	74.7	346112	3	US-09-949-016-13165	Sequence 13165, A	C 659	13.8	72.6	601	3	US-09-949-016-34576	Sequence 34576, A
C 587	14	73.7	462	3	US-09-270-767-2290	Sequence 2290, Ap	C 660	13.8	72.6	601	3	US-09-949-016-38553	Sequence 38553, A
C 588	14	73.7	462	3	US-09-270-767-17572	Sequence 17572, A	C 661	13.8	72.6	601	3	US-09-949-016-42831	Sequence 42831, A
C 589	14	73.7	497	3	US-09-328-475C-220	Sequence 220, App	C 662	13.8	72.6	601	3	US-09-949-016-45925	Sequence 45925, A
C 590	14	73.7	688	3	US-09-270-767-4835	Sequence 4835, Ap	C 663	13.8	72.6	601	3	US-09-949-016-46247	Sequence 46247, A
C 591	14	73.7	688	3	US-09-270-767-20117	Sequence 20117, A	C 664	13.8	72.6	601	3	US-09-949-016-46248	Sequence 46248, A
C 592	14	73.7	787	3	US-09-328-475C-336	Sequence 336, App	C 665	13.8	72.6	601	3	US-09-949-016-46249	Sequence 46249, A
C 593	14	73.7	831	3	US-09-328-475C-221	Sequence 221, App	C 666	13.8	72.6	601	3	US-09-949-016-48443	Sequence 48443, A
C 594	14	73.7	1236	3	US-09-540-236-1629	Sequence 1629, Ap	C 667	13.8	72.6	601	3	US-09-949-016-48444	Sequence 48444, A
C 595	14	73.7	1236	3	US-09-268-140-6	Sequence 6, Appli	C 668	13.8	72.6	601	3	US-09-949-016-48445	Sequence 48445, A
C 596	14	73.7	99629	3	US-09-596-002-37	Sequence 37, Appl	C 669	13.8	72.6	601	3	US-09-949-016-48716	Sequence 48716, A
C 597	13.8	72.6	187	3	US-09-513-999C-21586	Sequence 21586, A	C 670	13.8	72.6	601	3	US-09-949-016-48717	Sequence 48717, A
C 598	13.8	72.6	249	3	US-09-134-001C-114	Sequence 114, App	C 671	13.8	72.6	601	3	US-09-949-016-52063	Sequence 52063, A
C 599	13.8	72.6	265	3	US-09-513-999C-34761	Sequence 34761, A	C 672	13.8	72.6	601	3	US-09-949-016-54430	Sequence 54430, A
C 600	13.8	72.6	282	5	US-09-974-300-3743	Sequence 3743, Ap	C 673	13.8	72.6	601	3	US-09-949-016-55106	Sequence 55106, A
C 601	13.8	72.6	294	3	US-09-621-976-3324	Sequence 3324, Ap	C 674	13.8	72.6	601	3	US-09-949-016-55138	Sequence 55138, A
C 602	13.8	72.6	305	3	US-09-313-294A-6618	Sequence 6618, Ap	C 675	13.8	72.6	601	3	US-09-949-016-55636	Sequence 55636, A
C 603	13.8	72.6	330	3	US-09-702-705-1379	Sequence 1379, Ap	C 676	13.8	72.6	601	3	US-09-949-016-57018	Sequence 57018, A
C 604	13.8	72.6	330	3	US-09-736-457-1379	Sequence 1379, Ap	C 677	13.8	72.6	601	3	US-09-949-016-58037	Sequence 58037, A
C 605	13.8	72.6	330	3	US-09-614-124B-1379	Sequence 1379, Ap	C 678	13.8	72.6	601	3	US-09-949-016-58503	Sequence 58503, A
C 606	13.8	72.6	330	3	US-09-671-325-1379	Sequence 1379, Ap	C 679	13.8	72.6	601	3	US-09-949-016-61009	Sequence 61009, A
C 607	13.8	72.6	330	3	US-09-658-824-1379	Sequence 1379, Ap	C 680	13.8	72.6	601	3	US-09-949-016-66355	Sequence 66355, A

681	13.8	72.6	601	3	US-09-949-016-68074	Sequence 58074, A	754	13.8	72.6	905	3	US-09-774-639-85	Sequence 85, Appl
c 682	13.8	72.6	601	3	US-09-949-016-77543	Sequence 77543, A	c 755	13.8	72.6	915	3	US-09-248-796A-1167	Sequence 1167, Ap
c 683	13.8	72.6	601	3	US-09-949-016-93849	Sequence 93849, A	c 756	13.8	72.6	915	3	US-09-248-796A-4782	Sequence 4782, Ap
684	13.8	72.6	601	3	US-09-949-016-108134	Sequence 108134, A	c 757	13.8	72.6	930	5	US-10-770-127-57	Sequence 57, Appl
685	13.8	72.6	601	3	US-09-949-016-108236	Sequence 108236, A	c 758	13.8	72.6	930	5	US-10-770-127-61	Sequence 61, Appl
686	13.8	72.6	601	3	US-09-949-016-108237	Sequence 108237, A	c 759	13.8	72.6	930	5	US-10-770-127-63	Sequence 63, Appl
687	13.8	72.6	601	3	US-09-949-016-108296	Sequence 108296, A	c 760	13.8	72.6	930	5	US-10-770-127-182	Sequence 182, App
688	13.8	72.6	601	3	US-09-949-016-108297	Sequence 108297, A	c 761	13.8	72.6	930	5	US-10-770-127-190	Sequence 190, App
689	13.8	72.6	601	3	US-09-949-016-108326	Sequence 108326, A	c 762	13.8	72.6	930	5	US-10-770-127-192	Sequence 192, App
690	13.8	72.6	601	3	US-09-949-016-108327	Sequence 108327, A	c 763	13.8	72.6	960	5	US-10-770-127-52	Sequence 52, Appl
691	13.8	72.6	601	3	US-09-949-016-108328	Sequence 108328, A	c 764	13.8	72.6	999	3	US-09-248-796A-5959	Sequence 5959, Ap
c 692	13.8	72.6	601	3	US-09-949-016-111665	Sequence 111665, A	c 765	13.8	72.6	1083	3	US-09-134-001C-1520	Sequence 1520, Ap
c 693	13.8	72.6	601	3	US-09-949-016-111813	Sequence 111813, A	c 766	13.8	72.6	1119	4	US-09-605-703B-531	Sequence 531, App
c 694	13.8	72.6	601	3	US-09-949-016-111959	Sequence 111959, A	c 767	13.8	72.6	1172	3	US-09-016-434-1322	Sequence 1322, Ap
c 695	13.8	72.6	601	3	US-09-949-016-112104	Sequence 112104, A	c 768	13.8	72.6	1172	3	US-09-023-655-1320	Sequence 1320, Ap
c 696	13.8	72.6	601	3	US-09-949-016-113383	Sequence 113383, A	c 769	13.8	72.6	1172	3	US-09-917-254-44	Sequence 44, Appl
c 697	13.8	72.6	601	3	US-09-949-016-127123	Sequence 127123, A	c 770	13.8	72.6	1172	5	US-09-913-171A-10	Sequence 10, Appl
c 698	13.8	72.6	601	3	US-09-949-016-127124	Sequence 127124, A	c 771	13.8	72.6	1172	3	US-09-663-600A-143	Sequence 143, Appl
c 699	13.8	72.6	601	3	US-09-949-016-127125	Sequence 127125, A	c 772	13.8	72.6	1185	2	US-08-493-638-5	Sequence 5, Appli
c 700	13.8	72.6	601	3	US-09-949-016-133650	Sequence 133650, A	c 773	13.8	72.6	1245	3	US-09-614-221A-469	Sequence 469, App
c 701	13.8	72.6	601	3	US-09-949-016-133651	Sequence 133651, A	c 774	13.8	72.6	1259	3	US-09-270-767-31484	Sequence 31484, A
c 702	13.8	72.6	601	3	US-09-949-016-139254	Sequence 139254, A	c 775	13.8	72.6	1261	3	US-08-858-207A-151	Sequence 151, App
c 703	13.8	72.6	601	3	US-09-949-016-139255	Sequence 139255, A	c 776	13.8	72.6	1312	3	US-08-961-083-69	Sequence 69, Appl
c 704	13.8	72.6	601	3	US-09-949-016-139256	Sequence 139256, A	c 777	13.8	72.6	1312	3	US-09-536-784-69	Sequence 69, Appl
c 705	13.8	72.6	601	3	US-09-949-016-144764	Sequence 144764, A	c 778	13.8	72.6	1312	3	US-09-765-271-69	Sequence 69, Appl
c 706	13.8	72.6	601	3	US-09-949-016-144765	Sequence 144765, A	c 779	13.8	72.6	1312	3	US-09-765-272A-69	Sequence 69, Appl
c 707	13.8	72.6	601	3	US-09-949-016-146926	Sequence 146926, A	c 780	13.8	72.6	1410	3	US-09-328-352-2150	Sequence 2150, Ap
c 708	13.8	72.6	601	3	US-09-949-016-146927	Sequence 146927, A	c 781	13.8	72.6	1424	3	US-09-620-405B-464	Sequence 464, App
c 709	13.8	72.6	601	3	US-09-949-016-147316	Sequence 147316, A	c 782	13.8	72.6	1424	3	US-09-433-826B-464	Sequence 464, App
c 710	13.8	72.6	601	3	US-09-949-016-149205	Sequence 149205, A	c 783	13.8	72.6	1424	3	US-09-604-287A-464	Sequence 464, App
c 711	13.8	72.6	601	3	US-09-949-016-149206	Sequence 149206, A	c 784	13.8	72.6	1424	3	US-09-834-759-464	Sequence 464, App
c 712	13.8	72.6	601	3	US-09-949-016-150488	Sequence 150488, A	c 785	13.8	72.6	1424	3	US-09-590-751A-464	Sequence 464, App
c 713	13.8	72.6	601	3	US-09-949-016-150489	Sequence 150489, A	c 786	13.8	72.6	1424	3	US-09-551-621-464	Sequence 464, App
c 714	13.8	72.6	601	3	US-09-949-016-150503	Sequence 150503, A	c 787	13.8	72.6	1424	3	US-10-076-622-464	Sequence 464, App
c 715	13.8	72.6	601	3	US-09-949-016-152812	Sequence 152812, A	c 788	13.8	72.6	1424	4	US-10-124-805-464	Sequence 464, App
c 716	13.8	72.6	601	3	US-09-949-016-160355	Sequence 160355, A	c 789	13.8	72.6	1424	3	US-09-663-600A-49	Sequence 49, Appl
c 717	13.8	72.6	601	3	US-09-949-016-160356	Sequence 160356, A	c 790	13.8	72.6	1432	3	US-09-563-600A-50	Sequence 50, Appl
c 718	13.8	72.6	601	3	US-09-949-016-164435	Sequence 164435, A	c 791	13.8	72.6	1452	3	US-09-527-431-50	Sequence 50, Appl
c 719	13.8	72.6	601	3	US-09-949-016-164436	Sequence 164436, A	c 792	13.8	72.6	1452	3	US-09-446-861-50	Sequence 50, Appl
c 720	13.8	72.6	601	3	US-09-949-016-171192	Sequence 171192, A	c 793	13.8	72.6	1452	3	US-09-107-532A-587	Sequence 587, App
c 721	13.8	72.6	601	3	US-09-949-016-171322	Sequence 171322, A	c 794	13.8	72.6	1482	3	US-09-107-532A-762	Sequence 762, App
c 722	13.8	72.6	601	3	US-09-949-016-178132	Sequence 178132, A	c 795	13.8	72.6	1536	3	US-09-830-230A-588	Sequence 588, App
c 723	13.8	72.6	601	3	US-09-949-016-178452	Sequence 178452, A	c 796	13.8	72.6	1542	3	US-09-248-796A-11563	Sequence 11563, A
c 724	13.8	72.6	601	3	US-09-949-016-197046	Sequence 197046, A	c 797	13.8	72.6	1578	3	US-09-543-681A-1714	Sequence 1714, Ap
c 725	13.8	72.6	601	3	US-09-949-016-197666	Sequence 197666, A	c 798	13.8	72.6	1587	3	US-09-830-230A-587	Sequence 587, App
c 726	13.8	72.6	601	3	US-09-949-016-197667	Sequence 197667, A	c 799	13.8	72.6	1599	3	US-09-107-532A-2530	Sequence 31, Appl
c 727	13.8	72.6	601	3	US-09-949-016-199994	Sequence 199994, A	c 800	13.8	72.6	1661	3	US-09-166-350-31	Sequence 36, Appl
c 728	13.8	72.6	601	3	US-09-949-016-200068	Sequence 200068, A	c 801	13.8	72.6	1661	3	US-09-320-132-36	Sequence 1340, Ap
c 729	13.8	72.6	601	3	US-09-949-016-200190	Sequence 200190, A	c 802	13.8	72.6	1661	3	US-09-023-655-1340	Sequence 466, App
c 730	13.8	72.6	601	3	US-09-949-016-205133	Sequence 205133, A	c 803	13.8	72.6	1661	3	US-09-620-405B-466	Sequence 466, App
c 731	13.8	72.6	601	3	US-09-949-002-4265	Sequence 4265, Ap	c 804	13.8	72.6	1729	3	US-09-433-826B-466	Sequence 466, App
c 732	13.8	72.6	601	3	US-09-949-002-4266	Sequence 4266, Ap	c 805	13.8	72.6	1729	3	US-09-604-287A-466	Sequence 466, App
c 733	13.8	72.6	601	3	US-09-949-002-5165	Sequence 5165, Ap	c 806	13.8	72.6	1729	3	US-09-834-759-466	Sequence 466, App
c 734	13.8	72.6	601	3	US-09-949-002-9165	Sequence 9165, Ap	c 807	13.8	72.6	1729	3	US-09-590-751A-466	Sequence 466, App
c 735	13.8	72.6	639	3	US-08-956-171E-761	Sequence 761, App	c 808	13.8	72.6	1729	3	US-09-551-621-466	Sequence 466, App
c 736	13.8	72.6	639	3	US-08-781-986A-761	Sequence 761, App	c 809	13.8	72.6	1729	3	US-09-551-621A-466	Sequence 466, App
c 737	13.8	72.6	687	3	US-09-710-279-1861	Sequence 1861, Ap	c 810	13.8	72.6	1729	3	US-10-076-622-466	Sequence 466, App
c 738	13.8	72.6	722	3	US-09-107-532A-1520	Sequence 1520, Ap	c 811	13.8	72.6	1729	4	US-10-124-805-466	Sequence 466, App
c 739	13.8	72.6	722	3	US-09-621-976-17256	Sequence 17256, A	c 812	13.8	72.6	1750	3	US-08-949-770-1	Sequence 1, Appli
c 740	13.8	72.6	732	3	US-09-134-001C-1523	Sequence 1523, Ap	c 813	13.8	72.6	1779	3	US-09-489-039A-3610	Sequence 3610, Ap
c 741	13.8	72.6	751	3	US-08-956-171E-892	Sequence 892, App	c 814	13.8	72.6	1792	3	US-09-270-767-15162	Sequence 2, Appli
c 742	13.8	72.6	751	3	US-08-781-986A-892	Sequence 892, App	c 815	13.8	72.6	1813	3	US-09-558-338-2	Sequence 2, Appli
c 743	13.8	72.6	801	3	US-09-134-000C-3287	Sequence 3287, Ap	c 816	13.8	72.6	1861	3	US-10-876-841A-2	Sequence 2, Appli
c 744	13.8	72.6	810	3	US-09-583-110-328	Sequence 328, App	c 817	13.8	72.6	1861	3	US-09-769-787-203	Sequence 203, App
c 745	13.8	72.6	810	3	US-09-769-787-293	Sequence 293, App	c 818	13.8	72.6	1884	3	US-09-593-110-2476	Sequence 2476, Ap
c 746	13.8	72.6	811	3	US-09-270-767-8314	Sequence 8314, Ap	c 819	13.8	72.6	1884	3	US-09-769-787-203	Sequence 203, App
c 747	13.8	72.6	811	3	US-09-270-767-2356	Sequence 2356, A	c 820	13.8	72.6	1907	3	US-08-924-629C-46	Sequence 46, Appl
c 748	13.8	72.6	840	3	US-09-107-433-122	Sequence 122, App	c 821	13.8	72.6	1914	3	US-09-107-433-566	Sequence 566, App
c 749	13.8	72.6	885	5	US-10-770-127-193	Sequence 193, App	c 822	13.8	72.6	1914	3	US-09-107-433-1788	Sequence 1788, Ap
c 750	13.8	72.6	887	3	US-09-270-767-14919	Sequence 14919, A	c 823	13.8	72.6	1946	3	US-09-042-225-5	Sequence 5, Appli
c 751	13.8	72.6	900	5	US-10-770-127-184	Sequence 184, App	c 824	13.8	72.6	1953	3	US-10-076-622-549	Sequence 549, App
c 752	13.8	72.6	903	3	US-05-393-634-63	Sequence 63, Appl	c 825	13.8	72.6	1953	4	US-10-124-805-549	Sequence 549, App
c 753	13.8	72.6	903	5	US-10-770-127-29	Sequence 29, Appl	c 826	13.8	72.6	1970	4	US-10-094-749-1548	Sequence 1548, Ap

C 827	13.8	72.6	2028	4	US-10-094-749-374	Sequence 374, App	C 900	13.8	72.6	3681	3	US-09-590-751A-463	Sequence 463, App
C 828	13.8	72.6	2091	3	US-09-248-796A-5314	Sequence 5314, App	C 901	13.8	72.6	3681	3	US-09-551-621-463	Sequence 463, App
C 829	13.8	72.6	2099	3	US-09-107-532A-1551	Sequence 1551, App	C 902	13.8	72.6	3681	3	US-09-551-621A-463	Sequence 463, App
C 830	13.8	72.6	2121	3	US-09-973-278-831	Sequence 831, App	C 903	13.8	72.6	3681	3	US-10-076-622-463	Sequence 463, App
C 831	13.8	72.6	2183	2	US-07-993-073-1	Sequence 1, Appli	C 904	13.8	72.6	3681	4	US-10-124-805-463	Sequence 463, App
C 832	13.8	72.6	2193	3	US-09-398-865A-1	Sequence 1, Appli	C 905	13.8	72.6	3681	4	US-09-543-679A-2749	Sequence 2749, App
C 833	13.8	72.6	2217	3	US-09-710-714-1	Sequence 1, Appli	C 906	13.8	72.6	3681	5	US-09-543-679A-2748	Sequence 2748, App
C 834	13.8	72.6	2232	3	US-09-248-796A-2563	Sequence 2563, App	C 907	13.8	72.6	3681	5	US-09-543-679A-2748	Sequence 2748, App
C 835	13.8	72.6	2232	3	US-09-620-405B-491	Sequence 491, App	C 908	13.8	72.6	3681	3	US-09-949-002-5	Sequence 5, Appli
C 836	13.8	72.6	2232	3	US-09-834-759-491	Sequence 491, App	C 909	13.8	72.6	3681	3	US-09-949-002-133	Sequence 133, App
C 837	13.8	72.6	2232	3	US-10-076-622-491	Sequence 491, App	C 910	13.8	72.6	3681	3	US-09-620-405B-474	Sequence 474, App
C 838	13.8	72.6	2232	3	US-10-124-805-491	Sequence 491, App	C 911	13.8	72.6	3681	3	US-09-604-287A-474	Sequence 474, App
C 839	13.8	72.6	2232	4	US-08-674-887A-7	Sequence 7, Appli	C 912	13.8	72.6	3681	3	US-09-834-759-474	Sequence 474, App
C 840	13.8	72.6	2238	3	US-08-951-844-7	Sequence 7, Appli	C 913	13.8	72.6	3681	3	US-09-590-751A-474	Sequence 474, App
C 841	13.8	72.6	2238	3	US-09-412-347-7	Sequence 7, Appli	C 914	13.8	72.6	3681	3	US-09-551-621-474	Sequence 474, App
C 842	13.8	72.6	2241	3	US-09-743-237-1	Sequence 1, Appli	C 915	13.8	72.6	3681	3	US-09-551-621A-474	Sequence 474, App
C 843	13.8	72.6	2241	3	US-10-876-841A-1	Sequence 1, Appli	C 916	13.8	72.6	3681	4	US-10-076-622-474	Sequence 474, App
C 844	13.8	72.6	2286	3	US-08-956-171E-626	Sequence 626, App	C 917	13.8	72.6	3681	4	US-10-124-805-474	Sequence 474, App
C 845	13.8	72.6	2286	3	US-08-781-986A-626	Sequence 626, App	C 918	13.8	72.6	3681	3	US-09-601-198-50	Sequence 50, Appli
C 846	13.8	72.6	2319	3	US-09-590-991-7	Sequence 7, Appli	C 919	13.8	72.6	3681	2	US-08-396-479B-11	Sequence 11, Appli
C 847	13.8	72.6	2320	3	US-10-104-047-1823	Sequence 1823, App	C 920	13.8	72.6	3681	4	US-08-818-823-11	Sequence 11, Appli
C 848	13.8	72.6	2406	2	US-08-396-479B-7	Sequence 7, Appli	C 921	13.8	72.6	3681	4	US-09-949-016-1702	Sequence 1702, App
C 849	13.8	72.6	2406	2	US-08-818-823-7	Sequence 7, Appli	C 922	13.8	72.6	3681	4	US-03-880-107-2213	Sequence 2213, App
C 850	13.8	72.6	2416	3	US-09-201-936-41	Sequence 41, Appli	C 923	13.8	72.6	3681	7	PCT-US94-07297-40	Sequence 40, Appli
C 851	13.8	72.6	2416	3	US-09-011-356-41	Sequence 41, Appli	C 924	13.8	72.6	3681	4	US-10-076-622-566	Sequence 566, App
C 852	13.8	72.6	2416	3	US-09-672-717-228	Sequence 228, App	C 925	13.8	72.6	3681	4	US-10-124-805-566	Sequence 566, App
C 853	13.8	72.6	2416	3	US-09-201-932-41	Sequence 41, Appli	C 926	13.8	72.6	3681	3	US-09-620-312D-156	Sequence 156, App
C 854	13.8	72.6	2416	4	US-08-654-743-41	Sequence 41, Appli	C 927	13.8	72.6	3681	3	US-09-949-016-4228	Sequence 4228, App
C 855	13.8	72.6	2437	3	US-10-104-047-1159	Sequence 1159, App	C 928	13.8	72.6	3681	3	US-10-076-622-564	Sequence 564, App
C 856	13.8	72.6	2452	4	US-10-094-749-306	Sequence 306, App	C 929	13.8	72.6	3681	4	US-10-124-805-564	Sequence 564, App
C 857	13.8	72.6	2559	4	US-10-094-749-1159	Sequence 1159, App	C 930	13.8	72.6	3681	4	US-10-094-749-1094	Sequence 1094, App
C 858	13.8	72.6	2561	3	US-09-270-542-101	Sequence 101, App	C 931	13.8	72.6	3681	2	US-08-819-288-2	Sequence 2, Appli
C 859	13.8	72.6	2561	3	US-09-270-542-119	Sequence 119, App	C 932	13.8	72.6	3681	2	US-09-400-348-2	Sequence 2, Appli
C 860	13.8	72.6	2563	2	US-08-553-436A-7	Sequence 7, Appli	C 933	13.8	72.6	3681	2	US-08-261-822A-2	Sequence 2, Appli
C 861	13.8	72.6	2647	2	US-08-396-479B-9	Sequence 9, Appli	C 934	13.8	72.6	3681	7	PCT-US95-07744A-2	Sequence 2, Appli
C 862	13.8	72.6	2647	2	US-08-818-823-9	Sequence 9, Appli	C 935	13.8	72.6	3681	3	US-08-987-123-4	Sequence 4, Appli
C 863	13.8	72.6	2662	4	US-10-094-749-363	Sequence 363, App	C 936	13.8	72.6	3681	3	US-09-976-594-907	Sequence 907, App
C 864	13.8	72.6	2700	3	US-08-956-171E-511	Sequence 511, App	C 937	13.8	72.6	3681	2	US-08-261-822A-1	Sequence 1, Appli
C 865	13.8	72.6	2700	3	US-08-781-986A-511	Sequence 511, App	C 938	13.8	72.6	3681	7	PCT-US95-07744A-1	Sequence 1, Appli
C 866	13.8	72.6	2716	3	US-09-799-451-295	Sequence 295, App	C 939	13.8	72.6	3681	2	US-08-819-288-1	Sequence 1, Appli
C 867	13.8	72.6	2723	3	US-08-961-527-64	Sequence 64, Appli	C 940	13.8	72.6	3681	2	US-09-400-348-1	Sequence 1, Appli
C 868	13.8	72.6	2770	3	US-09-976-594-322	Sequence 322, App	C 941	13.8	72.6	3681	3	US-08-652-877-17	Sequence 17, Appli
C 869	13.8	72.6	2814	4	US-09-880-107-3419	Sequence 3419, App	C 942	13.8	72.6	3681	3	US-08-476-515A-17	Sequence 17, Appli
C 870	13.8	72.6	2832	3	US-08-476-515A-11	Sequence 11, Appli	C 943	13.8	72.6	3681	3	US-09-583-110-2613	Sequence 2613, App
C 871	13.8	72.6	2835	3	US-08-652-877-11	Sequence 11, Appli	C 944	13.8	72.6	3681	3	US-09-949-016-597	Sequence 597, App
C 872	13.8	72.6	2862	3	US-08-569-749-13	Sequence 13, Appli	C 945	13.8	72.6	3681	3	US-09-949-016-1533	Sequence 1533, App
C 873	13.8	72.6	2862	3	US-08-689-366-13	Sequence 13, Appli	C 946	13.8	72.6	3681	3	US-09-949-016-14734	Sequence 14734, A
C 874	13.8	72.6	2862	3	US-10-232-286-13	Sequence 13, Appli	C 947	13.8	72.6	3681	3	US-10-082-272-1	Sequence 1, Appli
C 875	13.8	72.6	2862	3	US-10-934-717-13	Sequence 13, Appli	C 948	13.8	72.6	3681	3	US-09-553-690-5	Sequence 5, Appli
C 876	13.8	72.6	2862	7	PCT-US96-12860-13	Sequence 13, Appli	C 949	13.8	72.6	3681	3	US-09-949-016-2318	Sequence 2318, App
C 877	13.8	72.6	2901	3	US-09-710-279-3425	Sequence 3425, App	C 950	13.8	72.6	3681	3	US-09-949-016-1063	Sequence 1063, App
C 878	13.8	72.6	3001	3	US-09-539-333D-123	Sequence 123, App	C 951	13.8	72.6	3681	3	US-08-961-527-75	Sequence 75, Appli
C 879	13.8	72.6	3045	4	US-10-076-622-548	Sequence 548, App	C 952	13.8	72.6	3681	3	US-09-949-016-16908	Sequence 16908, A
C 880	13.8	72.6	3045	4	US-10-124-805-548	Sequence 548, App	C 953	13.8	72.6	3681	2	US-08-392-625-16	Sequence 16, Appli
C 881	13.8	72.6	3052	3	US-09-710-279-3714	Sequence 3714, App	C 954	13.8	72.6	3681	2	US-08-466-961A-16	Sequence 16, Appli
C 882	13.8	72.6	3191	3	US-09-212-971-13	Sequence 13, Appli	C 955	13.8	72.6	3681	2	US-08-645-171B-17	Sequence 17, Appli
C 883	13.8	72.6	3151	3	US-08-800-929A-13	Sequence 13, Appli	C 956	13.8	72.6	3681	3	US-08-956-171E-38	Sequence 38, Appli
C 884	13.8	72.6	3151	3	US-09-617-053A-13	Sequence 13, Appli	C 957	13.8	72.6	3681	3	US-08-781-986A-37	Sequence 37, Appli
C 885	13.8	72.6	3164	3	US-09-976-594-320	Sequence 320, App	C 958	13.8	72.6	3681	3	US-08-956-171E-143	Sequence 143, App
C 886	13.8	72.6	3288	3	US-09-620-405B-490	Sequence 490, App	C 959	13.8	72.6	3681	3	US-08-781-986A-143	Sequence 143, App
C 887	13.8	72.6	3288	3	US-08-834-759-490	Sequence 490, App	C 960	13.8	72.6	3681	3	US-09-949-016-13091	Sequence 13091, A
C 888	13.8	72.6	3288	3	US-10-076-622-490	Sequence 490, App	C 961	13.8	72.6	3681	3	US-09-949-016-13092	Sequence 13092, A
C 889	13.8	72.6	3288	4	US-10-124-805-490	Sequence 490, App	C 962	13.8	72.6	3681	3	US-09-949-016-13093	Sequence 13093, A
C 890	13.8	72.6	3426	3	US-09-949-016-4547	Sequence 4547, App	C 963	13.8	72.6	3681	2	US-09-949-016-13094	Sequence 13094, A
C 891	13.8	72.6	3456	3	US-09-487-558B-133	Sequence 133, App	C 964	13.8	72.6	3681	3	US-09-949-016-13095	Sequence 13095, A
C 892	13.8	72.6	3583	5	US-09-543-679A-2750	Sequence 2750	C 965	13.8	72.6	3681	3	US-09-949-016-13320	Sequence 13320, A
C 893	13.8	72.6	3601	3	US-08-956-171E-6	Sequence 6, Appli	C 966	13.8	72.6	3681	3	US-09-949-016-13321	Sequence 13321, A
C 894	13.8	72.6	3601	3	US-08-781-986A-6	Sequence 6, Appli	C 967	13.8	72.6	3681	3	US-09-949-016-13322	Sequence 13322, A
C 895	13.8	72.6	3680	3	US-09-023-655-1477	Sequence 1477, App	C 968	13.8	72.6	3681	3		
C 896	13.8	72.6	3681	3	US-09-620-405B-463	Sequence 463, App	C 969	13.8	72.6	3681	3		
C 897	13.8	72.6	3681	3	US-09-433-826B-463	Sequence 463, App	C 970	13.8	72.6	3681	3		
C 898	13.8	72.6	3681	3	US-09-604-287A-463	Sequence 463, App	C 971	13.8	72.6	3681	3		
C 899	13.8	72.6	3681	3	US-09-834-759-463	Sequence 463, App	C 972	13.8	72.6	3681	3		

973 13.8 72.6 12383 3 US-09-949-016-13323 Sequence 13323, A  
974 13.8 72.6 12383 3 US-09-949-016-13324 Sequence 13324, A  
C 975 13.8 72.6 12785 3 US-09-553-690-1 Sequence 1, Appli  
C 976 13.8 72.6 13598 3 US-09-949-016-11856 Sequence 11856, A  
C 977 13.8 72.6 13599 3 US-09-949-016-15876 Sequence 15876, A  
C 978 13.8 72.6 13948 3 US-09-949-016-13023 Sequence 13023, A  
C 979 13.8 72.6 14042 3 US-08-652-877-85 Sequence 85, Appl  
C 980 13.8 72.6 14044 3 US-08-652-877-89 Sequence 89, Appl  
C 981 13.8 72.6 14080 3 US-08-652-877-87 Sequence 87, Appl  
C 982 13.8 72.6 14083 3 US-08-476-515A-83 Sequence 83, Appl  
C 983 13.8 72.6 14086 3 US-08-652-877-83 Sequence 83, Appl  
984 13.8 72.6 16134 3 US-09-949-016-13056 Sequence 13056, A  
985 13.8 72.6 16935 3 US-09-949-016-15534 Sequence 15534, A  
C 986 13.8 72.6 17069 5 US-09-543-679A-2752 Sequence 2752, Ap  
C 987 13.8 72.6 17132 3 US-09-949-016-15361 Sequence 15361, A  
C 988 13.8 72.6 19272 3 US-09-949-016-13462 Sequence 13462, A  
989 13.8 72.6 20099 3 US-09-949-016-13074 Sequence 13074, A  
990 13.8 72.6 20661 3 US-09-949-016-13326 Sequence 13326, A  
991 13.8 72.6 20661 3 US-09-949-016-13327 Sequence 13327, A  
992 13.8 72.6 20661 3 US-09-949-016-13328 Sequence 13328, A  
993 13.8 72.6 20661 3 US-09-949-016-13329 Sequence 13329, A  
994 13.8 72.6 20661 3 US-09-949-016-13330 Sequence 13330, A  
995 13.8 72.6 20661 3 US-09-949-016-13331 Sequence 13331, A  
996 13.8 72.6 20662 3 US-09-949-016-15156 Sequence 15156, A  
997 13.8 72.6 20681 3 US-09-949-016-14596 Sequence 14596, A  
998 13.8 72.6 20681 3 US-09-949-016-14597 Sequence 14597, A  
999 13.8 72.6 20681 3 US-09-949-016-14598 Sequence 14598, A  
1000 13.8 72.6 20681 3 US-09-949-016-14599 Sequence 14599, A

## ALIGNMENTS

RESULT 1  
US-09-949-016-29173/c  
; Sequence 29173, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 29173  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-29173

Query Match 100.0%; Score 19; DB 3; Length 601;  
Best Local Similarity 100.0%; Pred. No. 10;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCTTTG 19  
|||||  
Db 55 CTTTACTTCATAGTCTTTG 37

RESULT 2  
US-09-949-016-126957/c  
; Sequence 126957, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 126957  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-126957

Query Match 100.0%; Score 19; DB 3; Length 601;  
Best Local Similarity 100.0%; Pred. No. 10;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCTTTG 19  
|||||  
Db 55 CTTTACTTCATAGTCTTTG 37

RESULT 3  
US-09-188-930-249/c  
; Sequence 249, Application US/09188930A  
; Patent No. 6150502  
; GENERAL INFORMATION:  
; APPLICANT: Watson, James D.  
; APPLICANT: Strachan, Lorna  
; APPLICANT: Sleeman, Matthew  
; APPLICANT: Onrust, Rene  
; APPLICANT: Murison, James Greg  
; TITLE OF INVENTION: Compositions Isolated From Skin Cells  
; TITLE OF INVENTION: and Methods For Their Use  
; FILE REFERENCE: 11000.1011c1  
; CURRENT APPLICATION NUMBER: US/09/188,930A  
; CURRENT FILING DATE: 1998-11-09  
; NUMBER OF SEQ ID NOS: 348  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 249  
; LENGTH: 1212  
; TYPE: DNA  
; ORGANISM: Human  
US-09-188-930-249

Query Match 100.0%; Score 19; DB 3; Length 1212;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCTTTG 19  
|||||  
Db 1031 CTTTACTTCATAGTCTTTG 1013

RESULT 4  
US-09-312-283C-249/c  
; Sequence 249, Application US/09312283C  
; Patent No. 6573095  
; GENERAL INFORMATION:  
; APPLICANT: Watson, James D.  
; APPLICANT: Strachan, Lorna  
; APPLICANT: Sleeman, Matthew  
; APPLICANT: Onrust, Rene  
; APPLICANT: Murison, James G.  
; APPLICANT: Kumble, Krishanand D.  
; TITLE OF INVENTION: Compositions Isolated from Skin Cells  
; TITLE OF INVENTION: and Methods for Their Use

```
; FILE REFERENCE: 11000.1011c2
; CURRENT APPLICATION NUMBER: US/09/312,283C
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 425
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 249
; LENGTH: 1212
; TYPE: DNA
; ORGANISM: Mouse
US-09-312-283C-249

Query Match      100.0%; Score 19; DB 3; Length 1212;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19
   |||||
Db 1031 CTTTACTTCATAGTCTTTG 1013

RESULT 5
US-09-495-050A-223/c
; Sequence 223, Application US/09495050A
; Patent No. 6492505
; GENERAL INFORMATION:
; APPLICANT: Roopa, Reddy
; APPLICANT: Guegler, Karl, J.
; APPLICANT: Au-Young, Janice
; TITLE OF INVENTION: COMPOSITION FOR DETECTION OF GENES ENCODING MEMBRANE-ASSOCIATED
; FILE REFERENCE: PA-0013 US
; CURRENT APPLICATION NUMBER: US/09/495,050A
; CURRENT FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/118,318
; PRIOR FILING DATE: February 1, 1999
; NUMBER OF SEQ ID NOS: 305
; SOFTWARE: PERL Program
; SEQ ID NO 223
; LENGTH: 1228
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6492505 2681738CT1
US-09-495-050A-223

Query Match      100.0%; Score 19; DB 3; Length 1228;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19
   |||||
Db 582 CTTTACTTCATAGTCTTTG 564

RESULT 6
US-09-949-016-3613/c
; Sequence 3613, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
US-09-949-016-3613/c

Query Match      100.0%; Score 19; DB 3; Length 2289;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19
   |||||
Db 1158 CTTTACTTCATAGTCTTTG 1140

RESULT 7
US-09-949-016-572/c
; Sequence 572, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 572
; LENGTH: 2456
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-572

Query Match      100.0%; Score 19; DB 3; Length 2456;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19
   |||||
Db 1312 CTTTACTTCATAGTCTTTG 1294

RESULT 8
US-09-949-016-15355/c
; Sequence 15355, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
US-09-949-016-15355/c

Query Match      100.0%; Score 19; DB 3; Length 2608;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19
   |||||
Db 582 CTTTACTTCATAGTCTTTG 564
```

Query Match 100.0%; Score 19; DB 3; Length 26086;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19  
|||||  
Db 22954 CTTTACTTCATAGTCTTTG 22936

RESULT 9  
US-09-949-016-12314/c  
; Sequence 12314, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12314  
; LENGTH: 26238  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-12314

Query Match 100.0%; Score 19; DB 3; Length 26238;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19  
|||||  
Db 23108 CTTTACTTCATAGTCTTTG 23090

RESULT 10  
US-09-949-016-173650/c  
; Sequence 173650, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 173650  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-173650

Query Match 91.6%; Score 17.4; DB 3; Length 601;  
Best Local Similarity 94.7%; Pred. No. 58;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19  
|||||  
Db 221 CTTTACTTCATAGTCTTTG 203

RESULT 11  
US-09-949-016-16659  
; Sequence 16659, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 16659  
; LENGTH: 50950  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-16659

Query Match 91.6%; Score 17.4; DB 3; Length 50950;  
Best Local Similarity 94.7%; Pred. No. 89;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19  
|||||  
Db 21387 CTTTACTTCATAGTCTTTG 21405

RESULT 12  
US-09-513-999C-32761/c  
; Sequence 32761, Application US/09513999C  
; Patent No. 6783961  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Duclert, A.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.  
; Patent No. 6783961  
; FILE REFERENCE: 59.US2.REG  
; CURRENT APPLICATION NUMBER: US/09/513,999C  
; CURRENT FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/122,487  
; PRIOR FILING DATE: 1999-02-26  
; NUMBER OF SEQ ID NOS: 36681  
; SOFTWARE: Patent.pm  
; SEQ ID NO 32761  
; LENGTH: 407  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 384  
; OTHER INFORMATION: r=a or g  
US-09-513-999C-32761

Query Match 86.3%; Score 16.4; DB 3; Length 407;  
Best Local Similarity 94.4%; Pred. No. 1.7e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TTTTACTTCATAGTCTTTG 19  
|||||  
Db 40 TTTTACTTCATAGTCTTTG 23



```

; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16745
; LENGTH: 22615
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-16745

Query Match      86.3%; Score 16.4; DB 3; Length 22615;
Best Local Similarity 94.4%; Pred. No. 2.4e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      2 TTTACTTCATAGTCTTTG 19
      |||||
Db      18145 TTTACTTCATAGTCTTTG 18128

RESULT 16
US-09-949-016-17192
; Sequence 17192, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17192
; LENGTH: 107937
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc.feature
; LOCATION: (1)...(107937)
; OTHER INFORMATION: n = A,T,C or G
; US-09-949-016-17192

Query Match      86.3%; Score 16.4; DB 3; Length 107937;
Best Local Similarity 94.4%; Pred. No. 2.8e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 CTTTACTTCATAGTCTTT 18
      |||||
Db      42230 CTTTACTTCATAGTCTTT 42247

RESULT 17
US-09-949-016-17030
; Sequence 17030, Application US/09949016
; Patent No. 6812339

; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16805
; LENGTH: 15131
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-16805

Query Match      86.3%; Score 16.4; DB 3; Length 15131;
Best Local Similarity 94.4%; Pred. No. 2.3e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      2 TTTACTTCATAGTCTTTG 19
      |||||
Db      481 TTTACTTCATAGTCTTTG 464

RESULT 14
US-09-949-016-12148/c
; Sequence 12148, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12148
; LENGTH: 22614
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-12148

Query Match      86.3%; Score 16.4; DB 3; Length 22614;
Best Local Similarity 94.4%; Pred. No. 2.4e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      2 TTTACTTCATAGTCTTTG 19
      |||||
Db      18145 TTTACTTCATAGTCTTTG 18128

RESULT 15
US-09-949-016-16745/c
; Sequence 16745, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16745
; LENGTH: 22615
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-16745

Query Match      86.3%; Score 16.4; DB 3; Length 22615;
Best Local Similarity 94.4%; Pred. No. 2.4e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      2 TTTACTTCATAGTCTTTG 19
      |||||
Db      18145 TTTACTTCATAGTCTTTG 18128
```

```
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17030
; LENGTH: 139049
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(139049)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17030

Query Match      86.3%; Score 16.4; DB 3; Length 139049;
Best Local Similarity 94.4%; Pred. No. 2.8e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TTACTTCATAGTCTTG 19
Db 61326 TTACTTCATAGTCTTG 61343

RESULT 18
US-09-949-016-2159/c
; Sequence 2159, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2159
; LENGTH: 2363
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-2159

Query Match      84.2%; Score 16; DB 3; Length 2363;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TTACTTCATAGTCTT 17
Db 780 TTACTTCATAGTCTT 765

RESULT 19
US-09-949-016-13901/c
; Sequence 13901, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13901
; LENGTH: 22206
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(22206)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13901

Query Match      84.2%; Score 16; DB 3; Length 22206;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TTACTTCATAGTCTT 17
Db 9697 TTACTTCATAGTCTT 9682

RESULT 20
US-09-949-016-12273
; Sequence 12273, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12273
; LENGTH: 130563
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12273

Query Match      84.2%; Score 16; DB 3; Length 130563;
Best Local Similarity 100.0%; Pred. No. 4.3e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 TTACTTCATAGTCTTT 18
Db 111602 TTACTTCATAGTCTTT 111617

RESULT 21
US-09-949-016-16050
; Sequence 16050, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
```

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 16050  
; LENGTH: 131379  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-16050

Query Match 84.2%; Score 16; DB 3; Length 131379;  
Best Local Similarity 100.0%; Pred. No. 4.3e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 TTACTTCATAGTCTTT 18  
| | | | | | | | | | | | | | | | | | | | | |  
Db 111602 TTACTTCATAGTCTTT 111617

RESULT 22  
US-09-949-016-48463  
; Sequence 48463, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 48463  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-48463

Query Match 83.2%; Score 15.8; DB 3; Length 601;  
Best Local Similarity 89.5%; Pred. No. 3.3e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19  
| | | | | | | | | | | | | | | | | | | | | |  
Db 47 CTTTATTTCATATTCCTTG 65

RESULT 23  
US-09-949-016-48464  
; Sequence 48464, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14

; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 48464  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-48464

Query Match 83.2%; Score 15.8; DB 3; Length 601;  
Best Local Similarity 89.5%; Pred. No. 3.3e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19  
| | | | | | | | | | | | | | | | | | | | | |  
Db 219 CTTTATTTCATATTCCTTG 237

RESULT 24  
US-09-949-016-48465  
; Sequence 48465, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 48465  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-48465

Query Match 83.2%; Score 15.8; DB 3; Length 601;  
Best Local Similarity 89.5%; Pred. No. 3.3e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19  
| | | | | | | | | | | | | | | | | | | | | |  
Db 272 CTTTATTTCATATTCCTTG 290

RESULT 25  
US-09-949-016-48466  
; Sequence 48466, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498

```
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 48466
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-48466

Query Match      83.2%; Score 15.8; DB 3; Length 601;
Best Local Similarity 89.5%; Pred. No. 3.3e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCTTTG 19
Db 379 CTTTATTTCATATCTTTG 397

RESULT 26
US-09-949-016-48467
; Sequence 48467, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 48467
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-48467

Query Match      83.2%; Score 15.8; DB 3; Length 601;
Best Local Similarity 89.5%; Pred. No. 3.3e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCTTTG 19
Db 500 CTTTATTTCATATCTTTG 518

RESULT 27
US-09-949-016-48468
; Sequence 48468, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 48468
; LENGTH: 601
```

```
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-48468

Query Match      83.2%; Score 15.8; DB 3; Length 601;
Best Local Similarity 89.5%; Pred. No. 3.3e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCTTTG 19
Db 562 CTTTATTTCATATCTTTG 580

RESULT 28
US-09-949-016-121554/c
; Sequence 121554, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 121554
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-121554

Query Match      83.2%; Score 15.8; DB 3; Length 601;
Best Local Similarity 89.5%; Pred. No. 3.3e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCTTTG 19
Db 253 CTTTACTTCATAGTCTTTG 235

RESULT 29
US-09-270-767-12078
; Sequence 12078, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12078
; LENGTH: 1752
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-12078

Query Match      83.2%; Score 15.8; DB 3; Length 1752;
Best Local Similarity 89.5%; Pred. No. 3.6e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCTTTG 19
Db 379 CTCTACTTCATGGTCTTTG 397
```

```
RESULT 30
US-09-620-312D-1023/c
; Sequence 1023, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Fillinghast
; APPLICANT: Drmanac, Radje T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; FILE REFERENCE: Polypeptides
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: Pf-FL_genes Version 1.0
; SEQ ID NO 1023
; LENGTH: 2412
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (158)..(544)
US-09-620-312D-1023

Query Match      83.2%; Score 15.8; DB 3; Length 2412;
Best Local Similarity 89.5%; Pred. No. 3.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 CTTTACTTCATAGTCTTTG 19
Db      746 CTTTATTTCATAGCCTTG 728

RESULT 31
US-09-270-767-15279
; Sequence 15279, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15279
; LENGTH: 2833
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-15279

Query Match      83.2%; Score 15.8; DB 3; Length 2833;
Best Local Similarity 89.5%; Pred. No. 3.8e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 CTTTACTTCATAGTCTTTG 19
Db      746 CTTTATTTCATAGCCTTG 728

RESULT 32
US-09-949-016-15133/c
; Sequence 15133, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15133
; LENGTH: 27794
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15133

Query Match      83.2%; Score 15.8; DB 3; Length 27794;
Best Local Similarity 89.5%; Pred. No. 4.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 CTTTACTTCATAGTCTTTG 19
Db      22981 CTTTACTTCTAAGTCTTTG 22963

RESULT 33
US-09-949-016-15990/c
; Sequence 15990, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15990
; LENGTH: 48039
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(48039)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15990

Query Match      83.2%; Score 15.8; DB 3; Length 48039;
Best Local Similarity 89.5%; Pred. No. 4.9e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

QY 1 CTTACTTCATAGTCTTTG 19  
||| ||||| ||||| |||||  
Db 36983 CTTGACTTCATATCTTTG 36965

RESULT 34  
US-09-949-016-13175/c  
; Sequence 13175, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13175  
; LENGTH: 58768  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)...(58768)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-13175

Query Match 83.2%; Score 15.8; DB 3; Length 58768;  
Best Local Similarity 89.5%; Pred. No. 5e+02; Mismatches 2; Indels 0; Gaps 0;  
Matches 17; Conservative 0

QY 1 CTTACTTCATAGTCTTTG 19  
||| ||||| ||||| |||||  
Db 19852 CTTATTTCATATCTTTG 19834

RESULT 35  
US-09-949-016-12009  
; Sequence 12009, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12009  
; LENGTH: 76221  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-12009

Query Match 83.2%; Score 15.8; DB 3; Length 76221;  
Best Local Similarity 89.5%; Pred. No. 5.1e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTTACTTCATAGTCTTTG 19

Db 10151 CTTACTTCATAGTCATTG 10169  
||| ||||| ||||| ||||| |||||

RESULT 36  
US-09-949-016-17053  
; Sequence 17053, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 17053  
; LENGTH: 76221  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-17053

Query Match 83.2%; Score 15.8; DB 3; Length 76221;  
Best Local Similarity 89.5%; Pred. No. 5.1e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTTACTTCATAGTCTTTG 19  
||| ||||| ||||| ||||| |||||  
Db 10151 CTTACTTCATAGTCATTG 10169

RESULT 37  
US-09-949-016-16435/c  
; Sequence 16435, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 16435  
; LENGTH: 258775  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-16435

Query Match 83.2%; Score 15.8; DB 3; Length 258775;  
Best Local Similarity 89.5%; Pred. No. 5.6e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTTACTTCATAGTCTTTG 19  
||| ||||| ||||| ||||| |||||  
Db 65440 CTTATTACATAGTCTTTG 65422

RESULT 38

```
US-09-248-796A-154
; Sequence 154, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 154
; LENGTH: 279
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-154

Query Match      81.1%; Score 15.4; DB 3; Length 279;
Best Local Similarity 94.1%; Pred. No. 4.7e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      2 TTTACTTCATAGTCTTT 18
Db      156 TTTAGTTCATAGTCTTT 172

RESULT 39
US-09-397-787-141
; Sequence 141, Application US/09397787
; Patent No. 6468758
; GENERAL INFORMATION:
; APPLICANT: Benson, Darin R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR OVARIAN
; TITLE OF INVENTION: CANCER THERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.456C2
; CURRENT APPLICATION NUMBER: US/09/397,787
; CURRENT FILING DATE: 1999-09-16
; NUMBER OF SEQ ID NOS: 334
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 141
; LENGTH: 451
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(451)
; OTHER INFORMATION: n = A,T,C or G
US-09-397-787-141

Query Match      81.1%; Score 15.4; DB 3; Length 451;
Best Local Similarity 94.1%; Pred. No. 4.9e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      3 TTACTTCATAGTCTTTG 19
Db      404 TTACTTCAAGTCTTTG 420

RESULT 40
US-09-949-016-21055/c
; Sequence 21055, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 2000-04-14
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21055
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 2000-04-14
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21056
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-21056

Query Match      81.1%; Score 15.4; DB 3; Length 601;
Best Local Similarity 94.1%; Pred. No. 5.1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 CTTTACTTCATAGTCTT 17
Db      407 CTTTACTTCTTAGTCTT 391

RESULT 41
US-09-949-016-21056/c
; Sequence 21056, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21056
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-21056

Query Match      81.1%; Score 15.4; DB 3; Length 601;
Best Local Similarity 94.1%; Pred. No. 5.1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 CTTTACTTCATAGTCTT 17
Db      412 CTTTACTTCTTAGTCTT 396

RESULT 42
US-09-949-016-21057/c
; Sequence 21057, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
```

```
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21055
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-21055

Query Match      81.1%; Score 15.4; DB 3; Length 601;
Best Local Similarity 94.1%; Pred. No. 5.1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 CTTTACTTCATAGTCTT 17
Db      412 CTTTACTTCTTAGTCTT 396
```

```
US-09-949-016-21056/c
; Sequence 21056, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21056
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-21056

Query Match      81.1%; Score 15.4; DB 3; Length 601;
Best Local Similarity 94.1%; Pred. No. 5.1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 CTTTACTTCATAGTCTT 17
Db      407 CTTTACTTCTTAGTCTT 391
```

```
US-09-949-016-21057/c
; Sequence 21057, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21056
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-21056

Query Match      81.1%; Score 15.4; DB 3; Length 601;
Best Local Similarity 94.1%; Pred. No. 5.1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 CTTTACTTCATAGTCTT 17
Db      407 CTTTACTTCTTAGTCTT 391
```

```
US-09-949-016-21057/c
; Sequence 21057, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
```

```
US-09-949-016-21055/c
; Sequence 21055, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
```

```
US-09-949-016-21055/c
; Sequence 21055, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
```

```
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 21057
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-21057

Query Match      81.1%; Score 15.4; DB 3; Length 601;
Best Local Similarity 94.1%; Pred. No. 5.1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CTTTACTTCATAGTCTT 17
Db      288 CTTTACTTCATAGTCTT 272

RESULT 43
US-09-949-016-21058/c
; Sequence 21058, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 21058
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-21058

Query Match      81.1%; Score 15.4; DB 3; Length 601;
Best Local Similarity 94.1%; Pred. No. 5.1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CTTTACTTCATAGTCTT 17
Db      79 CTTTACTTCATAGTCTT 63

RESULT 44
US-09-949-016-60155/c
; Sequence 60155, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 60155
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-60155

Query Match      81.1%; Score 15.4; DB 3; Length 601;
Best Local Similarity 94.1%; Pred. No. 5.1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 TTTTACTTCATAGTCTT 18
Db      423 TTTTACTTCATAGTCTT 407

RESULT 45
US-09-949-016-77977/c
; Sequence 77977, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 77977
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-77977

Query Match      81.1%; Score 15.4; DB 3; Length 601;
Best Local Similarity 94.1%; Pred. No. 5.1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CTTTACTTCATAGTCTT 17
Db      412 CTTTACTTCATAGTCTT 396

RESULT 46
US-09-949-016-77978/c
; Sequence 77978, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 77978
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-77978
```



## US-09-949-016-77978

Query Match 81.1%; Score 15.4; DB 3; Length 601;  
Best Local Similarity 94.1%; Pred. No. 5.1e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTT 17  
|||||  
Db 407 CTTTACTTCATAGTCTTT 391

## RESULT 47

US-09-949-016-77979/c  
; Sequence 77979, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 77979  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-77979

Query Match 81.1%; Score 15.4; DB 3; Length 601;  
Best Local Similarity 94.1%; Pred. No. 5.1e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTT 17  
|||||  
Db 288 CTTTACTTCATAGTCTTT 272

## RESULT 48

US-09-949-016-77980/c  
; Sequence 77980, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 77980  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-77980

Query Match 81.1%; Score 15.4; DB 3; Length 601;  
Best Local Similarity 94.1%; Pred. No. 5.1e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTT 17  
|||||  
Db 79 CTTTACTTCATAGTCTTT 63

## RESULT 49

US-09-949-016-144809/c  
; Sequence 144809, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 144809  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-144809

Query Match 81.1%; Score 15.4; DB 3; Length 601;  
Best Local Similarity 94.1%; Pred. No. 5.1e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TTTACTTCATAGTCTTT 18  
|||||  
Db 162 TTTACTTTATAGTCTTT 146

## RESULT 50

US-09-949-016-162942/c  
; Sequence 162942, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 162942  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-162942

Query Match 81.1%; Score 15.4; DB 3; Length 601;  
Best Local Similarity 94.1%; Pred. No. 5.1e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 TTACTTCATAGTCTTTG 19  
|||||  
Db 508 TTACTTCATAGTCTTTG 492

```
RESULT 51
US-09-949-002-8840
; Sequence 8840, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; FILE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8840
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-002-8840

Query Match      81.1%; Score 15.4; DB 3; Length 601;
Best Local Similarity 94.1%; Pred. No. 5.1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3 TTACTTCATAGTCTTTG 19
      ||| ||||| ||||| |||
Db      268 TCACATTCATAGTCTTTG 284

RESULT 52
US-09-614-221A-382/c
; Sequence 382, Application US/09614221A
; Patent No. 6723837
; GENERAL INFORMATION:
; APPLICANT: Karunanandaa, Balasulojini
; APPLICANT: Yu, Jaehyuk
; APPLICANT: Kishore, Ganesh M.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED
; WITH STEROL SYNTHESIS AND METABOLISM
; FILE REFERENCE: 16516.075
; CURRENT APPLICATION NUMBER: US/09/614,221A
; CURRENT FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/142,981
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 626
; SEQ ID NO 382
; LENGTH: 2163
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-09-614-221A-382

Query Match      81.1%; Score 15.4; DB 3; Length 2163;
Best Local Similarity 94.1%; Pred. No. 5.7e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3 TTACTTCATAGTCTTTG 19
      ||| ||||| ||||| |||
Db      1662 TTCTTCATAGTCTTTG 1646

RESULT 53
US-10-104-047-659/c
; Sequence 659, Application US/10104047
; Patent No. 6943241
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. 6943241el full length cdna
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
```

```
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 659
; LENGTH: 2397
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-047-659

Query Match      81.1%; Score 15.4; DB 3; Length 2397;
Best Local Similarity 94.1%; Pred. No. 5.8e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3 TTACTTCATAGTCTTTG 19
      ||| ||||| ||||| |||
Db      1773 TTACTTCACAGTCTTTG 1757

RESULT 54
US-08-956-171E-295/c
; Sequence 295, Application US/08956171E
; Patent No. 6593114
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; APPLICANT: Gil H. Choi
; APPLICANT: Patrick S. Dillon
; APPLICANT: Craig A. Rosen
; APPLICANT: Steven C. Barash
; APPLICANT: Michael R. Fannon
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5256
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/956,171E
; FILING DATE: 20-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/009,861
; FILING DATE: January 5, 1996
; APPLICATION NUMBER: 08/781,986
; FILING DATE: January 3, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mark J. Hyman
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB248P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (240) 314-1224
; TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 295:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2817 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 295:
US-08-956-171E-295

Query Match      81.1%; Score 15.4; DB 3; Length 2817;
Best Local Similarity 94.1%; Pred. No. 5.9e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

Qy 2 TTTACTTCATAGCTCTT 18  
|||||  
Db 2446 TTTACTTCATAGCCTTT 2430

## RESULT 55

US-08-781-986A-295/c  
; Sequence 295, Application US/08781986A  
; Patent No. 6737248  
; GENERAL INFORMATION:  
; APPLICANT: Charles Kunsch  
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
; NUMBER OF SEQUENCES: 5255  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/781,986A  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Benson, Bob  
; REGISTRATION NUMBER: 30,446  
; REFERENCE/DOCKET NUMBER: P8248PP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 295:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2817 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
US-08-781-986A-295

Query Match 81.1%; Score 15.4; DB 3; Length 2817;  
Best Local Similarity 94.1%; Pred. No. 5.9e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TTTACTTCATAGCTCTT 18  
|||||  
Db 2446 TTTACTTCATAGCCTTT 2430

## RESULT 56

US-09-248-528-5  
; Sequence 5, Application US/09248528C  
; Patent No. 6153415  
; GENERAL INFORMATION:  
; APPLICANT: Oriel, Patrick J  
; APPLICANT: Padmakumar, Rugmini  
; APPLICANT: Kim, Sang H  
; TITLE OF INVENTION: Method for Producing Amide Compounds Using a Nitrile  
; TITLE OF INVENTION: Hydratase from a Thermophilic Bacillus  
; FILE REFERENCE: MSU 4.1-401  
; CURRENT APPLICATION NUMBER: US/09/248,528C  
; CURRENT FILING DATE: 1999-02-10  
; EARLIER APPLICATION NUMBER: 60/083,485  
; EARLIER FILING DATE: 1998-04-29  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 5  
; LENGTH: 3299  
; TYPE: DNA  
; ORGANISM: BR449  
; FEATURE:  
; NAME/KEY: gene  
; LOCATION: (432)..(1475)  
; OTHER INFORMATION: Amidase gene  
; FEATURE:  
; NAME/KEY: gene  
; LOCATION: (1606)..(2292)  
; OTHER INFORMATION: Nitrile hydratase beta subunit  
; FEATURE:  
; NAME/KEY: gene  
; LOCATION: (2321)..(2962)  
; OTHER INFORMATION: Nitrile hydratase alpha subunit  
; FEATURE:  
; NAME/KEY: gene  
; LOCATION: (2981)..(3283)  
; OTHER INFORMATION: ORF 1  
US-09-248-528-5

Query Match 81.1%; Score 15.4; DB 3; Length 3299;  
Best Local Similarity 94.1%; Pred. No. 5.9e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGCTCTT 17  
|||||  
Db 193 CTTTACTTCATAGCTCTT 209

## RESULT 57

US-09-549-108-5  
; Sequence 5, Application US/09549108  
; Patent No. 6214603  
; GENERAL INFORMATION:  
; APPLICANT: Oriel, Patrick J  
; APPLICANT: Padmakumar, Rugmini  
; APPLICANT: Kim, Sang H  
; TITLE OF INVENTION: Method for Producing Amide Compounds Using a Nitrile  
; TITLE OF INVENTION: Hydratase from a Thermophilic Bacillus  
; FILE REFERENCE: MSU 4.1-486  
; CURRENT APPLICATION NUMBER: US/09/549,108  
; CURRENT FILING DATE: 2000-04-13  
; PRIOR APPLICATION NUMBER: 60/083,485  
; PRIOR FILING DATE: 1998-04-29  
; PRIOR APPLICATION NUMBER: 09/248,528  
; PRIOR FILING DATE: 1999-02-10  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 5  
; LENGTH: 3299  
; TYPE: DNA  
; ORGANISM: BR449  
; FEATURE:  
; NAME/KEY: gene  
; LOCATION: (432)..(1475)  
; OTHER INFORMATION: Amidase gene  
; NAME/KEY: gene  
; LOCATION: (1606)..(2292)  
; OTHER INFORMATION: Nitrile hydratase beta subunit  
; NAME/KEY: gene  
; LOCATION: (2321)..(2962)  
; OTHER INFORMATION: Nitrile hydratase alpha subunit  
; NAME/KEY: gene  
; LOCATION: (2981)..(3283)  
; OTHER INFORMATION: ORF 1  
US-09-549-108-5

Query Match 81.1%; Score 15.4; DB 3; Length 3299;  
Best Local Similarity 94.1%; Pred. No. 5.9e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY 1 CTTTACTTCATAGTCCTT 17
||||| |||||
Db 193 CTTTACTTCTTAGTCCTT 209

RESULT 58
US-09-549-111-5
; Sequence 5, Application US/09549111
; Patent No. 6228633
; GENERAL INFORMATION:
; APPLICANT: Oriol, Patrick J
; APPLICANT: Padmakumar, Rugmini
; APPLICANT: Kim, Sang H
; TITLE OF INVENTION: Method for Producing Amide Compounds Using a Nitrile
; TITLE OF INVENTION: Hydratase from a Thermophilic Bacillus
; FILE REFERENCE: MSU 4.1-489
; CURRENT APPLICATION NUMBER: US/09/549,111
; CURRENT FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: 60/083,485
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 09/248,528
; PRIOR FILING DATE: 1999-02-10
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 3299
; TYPE: DNA
; ORGANISM: BR449
; FEATURE:
; NAME/KEY: gene
; LOCATION: (432)..(1475)
; OTHER INFORMATION: Amidase gene
; NAME/KEY: gene
; LOCATION: (1606)..(2292)
; OTHER INFORMATION: Nitrile hydratase beta subunit
; NAME/KEY: gene
; LOCATION: (2321)..(2962)
; OTHER INFORMATION: Nitrile hydratase alpha subunit
; NAME/KEY: gene
; LOCATION: (2981)..(3283)
; OTHER INFORMATION: ORF 1
US-09-549-111-5

Query Match 81.1%; Score 15.4; DB 3; Length 3299;
Best Local Similarity 94.1%; Pred. No. 5.9e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCCTT 17
||||| |||||
Db 193 CTTTACTTCTTAGTCCTT 209

RESULT 60
US-09-550-394-5
; Sequence 5, Application US/09550394
; Patent No. 6287828
; GENERAL INFORMATION:
; APPLICANT: Oriol, Patrick J
; APPLICANT: Padmakumar, Rugmini
; APPLICANT: Kim, Sang H
; TITLE OF INVENTION: Method for Producing Amide Compounds Using a Nitrile
; TITLE OF INVENTION: Hydratase from a Thermophilic Bacillus
; FILE REFERENCE: MSU 4.1-488
; CURRENT APPLICATION NUMBER: US/09/550,394
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/083,485
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 09/248,528
; PRIOR FILING DATE: 1999-02-10
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 3299
; TYPE: DNA
; ORGANISM: BR449
; FEATURE:
; NAME/KEY: gene
; LOCATION: (432)..(1475)
; OTHER INFORMATION: Amidase gene
; NAME/KEY: gene
; LOCATION: (1606)..(2292)
; OTHER INFORMATION: Nitrile hydratase beta subunit
; NAME/KEY: gene
; LOCATION: (2321)..(2962)
; OTHER INFORMATION: Nitrile hydratase alpha subunit
; NAME/KEY: gene
; LOCATION: (2981)..(3283)
; OTHER INFORMATION: ORF 1
US-09-550-394-5

Query Match 81.1%; Score 15.4; DB 3; Length 3299;
Best Local Similarity 94.1%; Pred. No. 5.9e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCCTT 17
||||| |||||
Db 193 CTTTACTTCTTAGTCCTT 209

RESULT 59
US-09-549-106-5
; Sequence 5, Application US/09549106
; Patent No. 6242242
; GENERAL INFORMATION:
; APPLICANT: Oriol, Patrick J
; APPLICANT: Padmakumar, Rugmini
; APPLICANT: Kim, Sang H
; TITLE OF INVENTION: Method for Producing Amide Compounds Using a Nitrile
; TITLE OF INVENTION: Hydratase from a Thermophilic Bacillus
; FILE REFERENCE: MSU 4.1-487
; CURRENT APPLICATION NUMBER: US/09/549,106
; CURRENT FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: 60/083,485
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 09/248,528
; PRIOR FILING DATE: 1999-02-10
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 3299
```

RESULT 61  
US-09-354-147C-6  
; Sequence 6, Application US/09354147C  
; Patent No. 6573067  
; GENERAL INFORMATION:  
; APPLICANT: Dib-Hajj, Sulayman  
; APPLICANT: Waxman, Stephen G.  
; TITLE OF INVENTION: Modulation of Sodium Channels in Dorsal Root Ganglia  
; FILE REFERENCE: 44574-5004-01-US  
; CURRENT APPLICATION NUMBER: US/09/354,147C  
; CURRENT FILING DATE: 1999-07-16  
; PRIOR APPLICATION NUMBER: US 60/072,990  
; PRIOR FILING DATE: 1998-01-29  
; PRIOR APPLICATION NUMBER: US 60/109,402  
; PRIOR FILING DATE: 1998-11-20  
; PRIOR APPLICATION NUMBER: PCT/US99/02008  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 3701  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(3699)  
; OTHER INFORMATION: partial human NaN cDNA sequence  
; NAME/KEY: unsure  
; LOCATION: (922)  
; OTHER INFORMATION: Y = c or t. Xaa at amino acid position 308 is  
; OTHER INFORMATION: Leu.  
US-09-354-147C-6

Query Match 81.1%; Score 15.4; DB 3; Length 3701;  
Best Local Similarity 94.1%; Pred. No. 6e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
Qy 2 TTTACTTCATAGTCTTT 18  
|||||  
Db 3260 TTTACTTCGTAAGTCTTT 3276

RESULT 62  
US-08-927-939A-30/c  
; Sequence 30, Application US/08927939A  
; Patent No. 6989435  
; GENERAL INFORMATION:  
; APPLICANT: Grainger, David J.  
; APPLICANT: Tatallick, Lauren Marie  
; TITLE OF INVENTION: Compounds and methods to inhibit or  
; TITLE OF INVENTION: augment an inflammatory response.  
; FILE REFERENCE: 1543.001US1  
; CURRENT APPLICATION NUMBER: US/08/927,939A  
; CURRENT FILING DATE: 1997-11-09  
; NUMBER OF SEQ ID NOS: 85  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 30  
; LENGTH: 4259  
; TYPE: DNA  
; ORGANISM: Mus musculus  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (2070)...(2130)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (2669)...(2795)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (2990)...(3079)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (3491)...(3506)  
US-08-927-939A-30

Query Match 81.1%; Score 15.4; DB 4; Length 4259;  
Best Local Similarity 94.1%; Pred. No. 6.1e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
Qy 2 TTTACTTCATAGTCTTT 18  
|||||  
Db 1612 TTTACTTCATAGGCTTT 1596

RESULT 63  
US-09-354-147C-41  
; Sequence 41, Application US/09354147C  
; Patent No. 6573067  
; GENERAL INFORMATION:  
; APPLICANT: Dib-Hajj, Sulayman  
; APPLICANT: Waxman, Stephen G.  
; TITLE OF INVENTION: Modulation of Sodium Channels in Dorsal Root Ganglia  
; FILE REFERENCE: 44574-5004-01-US  
; CURRENT APPLICATION NUMBER: US/09/354,147C  
; CURRENT FILING DATE: 1999-07-16  
; PRIOR APPLICATION NUMBER: US 60/072,990  
; PRIOR FILING DATE: 1998-01-29  
; PRIOR APPLICATION NUMBER: US 60/109,402  
; PRIOR FILING DATE: 1998-11-20  
; PRIOR APPLICATION NUMBER: PCT/US99/02008  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 41  
; LENGTH: 5860  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (31)..(5403)  
; OTHER INFORMATION: full length cDNA sequence for human NaN  
US-09-354-147C-41

Query Match 81.1%; Score 15.4; DB 3; Length 5860;  
Best Local Similarity 94.1%; Pred. No. 6.3e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
Qy 2 TTTACTTCATAGTCTTT 18  
|||||  
Db 3887 TTTACTTCGTAAGTCTTT 3903

RESULT 64  
US-09-220-132-38  
; Sequence 38, Application US/09220132  
; Patent No. 6506607  
; GENERAL INFORMATION:  
; APPLICANT: Shyjan, Andrew W.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE IDENTIFICATION AND ASSESSMENT  
; TITLE OF INVENTION: OF PROSTATE CANCER THERAPIES AND THE DIAGNOSIS OF PROSTATE CANCER  
; FILE REFERENCE: 07334-074001  
; CURRENT APPLICATION NUMBER: US/09/220,132  
; CURRENT FILING DATE: 1998-12-23  
; PRIOR APPLICATION NUMBER: US 60/079,303  
; PRIOR FILING DATE: 1998-03-25  
; PRIOR APPLICATION NUMBER: US 60/068,821  
; PRIOR FILING DATE: 1997-12-24  
; NUMBER OF SEQ ID NOS: 191  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 38  
; LENGTH: 7679  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-220-132-38

Query Match 81.1%; Score 15.4; DB 3; Length 7679;  
Best Local Similarity 94.1%; Pred. No. 6.4e+02;

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 TTACTTCATAGTCTTTG 19  
|||||  
Db 2839 TTACTTCAAAGTCTTTG 2855

RESULT 65  
US-09-023-655-1289  
; Sequence 1289, Application US/09023655  
; Patent No. 6607879  
; GENERAL INFORMATION:  
; APPLICANT: Cocks, Benjamin G.  
; APPLICANT: Susan G. Stuart  
; APPLICANT: Jeffrey J. Seilhamer  
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE  
; TITLE OF INVENTION: EXPRESSION  
; NUMBER OF SEQUENCES: 1508  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
; STREET: 3174 PORTER DRIVE  
; CITY: PALO ALTO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/023,655  
; FILING DATE: HERewith  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Zeller, Karen J.  
; REGISTRATION NUMBER: 37,071  
; REFERENCE/DOCKET NUMBER: PA-0001 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 855-0555  
; TELEFAX: (650) 845-4166  
; INFORMATION FOR SEQ ID NO: 1289:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7680 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: GENBANK  
; CLONE: g31396  
US-09-023-655-1289

Query Match 81.1%; Score 15.4; DB 3; Length 7680;  
Best Local Similarity 94.1%; Pred. No. 6.4e+02;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 TTACTTCATAGTCTTTG 19  
|||||  
Db 2839 TTACTTCAAAGTCTTTG 2855

RESULT 66  
US-09-543-679A-2698  
; Sequence 2698, Application US/09543679A  
; Patent No. 7034007  
; GENERAL INFORMATION:  
; APPLICANT: NYCE, Jonathan W.  
; TITLE OF INVENTION: LOW ADENOSINE ANTI-SENSE OLIGONUCLEOTIDE,  
; COMPOSITIONS, KIT & METHOD FOR TREATMENT

; OF AIRWAY DISORDERS ASSOCIATED WITH  
; BRONCHOCONSTRICTION, LUNG INFLAMMATION,  
; NUMBER OF SEQUENCES: 3111  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: EPIGENESIS PHARMACEUTICALS, INC.  
; STREET: 7 Clarke Drive  
; CITY: Cranbury  
; STATE: NJ  
; COUNTRY: USA  
; ZIP: 08512  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-R  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: N/A  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/543,679A  
; FILING DATE: 13-Apr-2000  
; CLASSIFICATION: UNKNOWN  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/127,958  
; FILING DATE: 1998-08-03  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Amzel, Viviana  
; REGISTRATION NUMBER: 30,930  
; REFERENCE/DOCKET NUMBER: EPI-0067191b  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 609-409-3035  
; TELEFAX: 413-254-9245  
; TELEX: <Unknown>  
; INFORMATION FOR SEQ ID NO: 2698:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7680 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; SEQUENCE DESCRIPTION: SEQ ID NO: 2698  
; US-09-543-679A-2698  
Query Match 81.1%; Score 15.4; DB 5; Length 7680;  
Best Local Similarity 94.1%; Pred. No. 6.4e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 3 TTACTTCATAGTCTTTG 19  
|||||  
Db 2839 TTACTTCAAAGTCTTTG 2855  
RESULT 67  
PCT-US95-09819-6  
; Sequence 6, Application PC/TUS9509819  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: FIBRONECTIN FIBRIN-BINDING PEPTIDES, DNA  
; TITLE OF INVENTION: CODING THEREFOR AND USES THEREOF  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/09819  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/283,857

;; FILING DATE: 01-AUG-1994  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: BROWDY, Roger L.  
;; REGISTRATION NUMBER: 25,618  
;; REFERENCE/DOCKET NUMBER: GOLD-1A PCT  
;; TELEPHONE: 202-628-5197  
;; TELEFAX: 202-737-3528  
;; TELEX: 248633  
;; INFORMATION FOR SEQ ID NO: 6:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 7680 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: CDNA  
PCT-US95-09819-6

Query Match 81.1%; Score 15.4; DB 7; Length 7680;  
Best Local Similarity 94.1%; Pred. No. 6.4e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 TTACTTCATAGTCTTTG 19  
Db 2839 TTACTTCAAGTCTTTG 2855

RESULT 68  
US-08-259-569-16  
; Sequence 16, Application US/08259569  
; Patent No. 5679320  
; GENERAL INFORMATION:  
; APPLICANT: Vogel, Tikva  
; APPLICANT: Levanon, Avigdor  
; APPLICANT: Werber, Moshe  
; APPLICANT: Guy, Rachel  
; APPLICANT: Panet, Amos  
; APPLICANT: Hartman, Jacob  
; APPLICANT: Shaked, Hadassa  
; TITLE OF INVENTION: FIBRIN BINDING DOMAIN POLYPEPTIDES AND  
; NUMBER OF SEQUENCES: 39  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooper & Dunham  
; STREET: 30 Rockefeller Plaza  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10112  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/259,569  
; FILING DATE:  
; CLASSIFICATION: 436  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 703,842  
; FILING DATE: 21-MAY-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: White, John P.  
; REGISTRATION NUMBER: 28,678  
; REFERENCE/DOCKET NUMBER: 25775-D/JPW/EAB  
; TELEPHONE: 212-664-0525  
; TELEFAX: 212-664-0525  
; TELEX: 422523 COOP UI  
; INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7705 base pairs  
; TYPE: nucleic acid

;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: CDNA  
;; HYPOTHETICAL: NO  
;; ANTI-SENSE: NO  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: 5..7681  
US-08-259-569-16

Query Match 81.1%; Score 15.4; DB 2; Length 7705;  
Best Local Similarity 94.1%; Pred. No. 6.4e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 TTACTTCATAGTCTTTG 19  
Db 2839 TTACTTCAAGTCTTTG 2855

RESULT 69  
US-08-826-885-16  
; Sequence 16, Application US/08826885  
; Patent No. 5869616  
; GENERAL INFORMATION:  
; APPLICANT: Vogel, Tikva  
; APPLICANT: Levanon, Avigdor  
; APPLICANT: Werber, Moshe  
; APPLICANT: Guy, Rachel  
; APPLICANT: Panet, Amos  
; APPLICANT: Hartman, Jacob  
; APPLICANT: Shaked, Hadassa  
; TITLE OF INVENTION: FIBRIN BINDING DOMAIN POLYPEPTIDES AND  
; NUMBER OF SEQUENCES: 39  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooper & Dunham  
; STREET: 30 Rockefeller Plaza  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10112  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/826,885  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/703,842  
; FILING DATE: 21-MAY-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: White, John P.  
; REGISTRATION NUMBER: 28,678  
; REFERENCE/DOCKET NUMBER: 25775-D/JPW/EAB  
; TELEPHONE: 212-664-0525  
; TELEFAX: 212-664-0525  
; TELEX: 422523 COOP UI  
; INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7705 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: CDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 5..7681





Qy 3 TTACTTCATAGTCTTTG 19  
|||||  
Db 2924 TTACTTCAAGTCTTTG 2940

RESULT 73  
US-09-566-921-135  
; Sequence 135, Application US/09566921  
; Patent No. 6682888  
; GENERAL INFORMATION:  
; APPLICANT: Loring, Jeanne F.  
; APPLICANT: Tingley, Debora W.  
; APPLICANT: Edwards, Carla M.  
; TITLE OF INVENTION: GENES EXPRESSED IN ALZHEIMER'S DISEASE  
; FILE REFERENCE: PA-0024 US  
; CURRENT APPLICATION NUMBER: US/09/566,921  
; CURRENT FILING DATE: 2000-05-05  
; NUMBER OF SEQ ID NOS: 138  
; SOFTWARE: PERL Program  
; SEQ ID NO 135  
; LENGTH: 8044  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. 6682888 427813.14  
US-09-566-921-135

Query Match 81.1%; Score 15.4; DB 3; Length 8044;  
Best Local Similarity 94.1%; Pred. No. 6.5e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 TTACTTCATAGTCTTTG 19  
|||||  
Db 3187 TTACTTCAAGTCTTTG 3203

RESULT 74  
US-09-543-679A-2699  
; Sequence 2699, Application US/09543679A  
; Patent No. 7034007  
; GENERAL INFORMATION:  
; APPLICANT: NYCE, Jonathan W.  
; TITLE OF INVENTION: LOW ADENOSINE ANTI-SENSE OLIGONUCLEOTIDE,  
; COMPOSITIONS, KIT & METHOD FOR TREATMENT  
; OF AIRWAY DISORDERS ASSOCIATED WITH  
; BRONCHOCONSTRICTION, LUNG INFLAMMATION,  
; NUMBER OF SEQUENCES: 3111  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: EPIGENESIS PHARMACEUTICALS, INC.  
; STREET: 7 Clarke Drive  
; CITY: Cranbury  
; STATE: NJ  
; COUNTRY: USA  
; ZIP: 08512  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-R  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: N/A  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/543,679A  
; FILING DATE: 13-Apr-2000  
; CLASSIFICATION: UNKNOWN  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/127,958  
; FILING DATE: 1998-08-03  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Anzel, Viviana  
; REGISTRATION NUMBER: 30,930  
; REFERENCE/DOCKET NUMBER: EPI-0067191b  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 609-409-3035

; TELEFAX: 413-254-9245  
; TELEX: <Unknown>  
; INFORMATION FOR SEQ ID NO: 2699:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 14740 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; SEQUENCE DESCRIPTION: SEQ ID NO: 2699  
US-09-543-679A-2699  
  
Query Match 81.1%; Score 15.4; DB 5; Length 14740;  
Best Local Similarity 94.1%; Pred. No. 6.8e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
Qy 3 TTACTTCATAGTCTTTG 19  
|||||  
Db 9899 TTACTTCAAGTCTTTG 9915

RESULT 75  
US-09-949-016-16147  
; Sequence 16147, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 16147  
; LENGTH: 39937  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-16147

Query Match 81.1%; Score 15.4; DB 3; Length 39937;  
Best Local Similarity 94.1%; Pred. No. 7.5e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTT 17  
|||||  
Db 36637 CTTTACTTCATAGTCTT 36653

RESULT 76  
US-09-949-016-11904/C  
; Sequence 11904, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 11904  
; LENGTH: 40624  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-11904

Query Match 81.1%; Score 15.4; DB 3; Length 40624;  
Best Local Similarity 94.1%; Pred. No. 7.5e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGCTTT 17  
|||||

Db 33515 CTTTACTTCATAGCTTT 33499

## RESULT 77

US-09-949-016-14016/c  
; Sequence 14016, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14016  
; LENGTH: 40625  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-14016

Query Match 81.1%; Score 15.4; DB 3; Length 40625;  
Best Local Similarity 94.1%; Pred. No. 7.5e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGCTTT 17  
|||||

Db 33515 CTTTACTTCATAGCTTT 33499

## RESULT 78

US-09-949-016-15836  
; Sequence 15836, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 15836  
; LENGTH: 45587  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-15836

Query Match 81.1%; Score 15.4; DB 3; Length 151256;  
Best Local Similarity 94.1%; Pred. No. 8.2e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 TTACTTCATAGTCTTTG 19  
|||||

Db 63211 TTCTTCATAGTCTTTG 63227

## RESULT 80

US-09-949-016-13242  
; Sequence 13242, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13242  
; LENGTH: 151261  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature

```

; LOCATION: (1)...(151261)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13242

Query Match      81.1%; Score 15.4; DB 3; Length 151261;
Best Local Similarity 94.1%; Pred. No. 8.2e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 TTACTTCATAGTCTTTG 19
||| ||||| ||||| |||||
Db 63211 TTCCTTCATAGTCTTTG 63227

RESULT 81
US-09-949-002-574/c
; Sequence 574, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 574
; LENGTH: 209631
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(209631)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-002-574

Query Match      81.1%; Score 15.4; DB 3; Length 209631;
Best Local Similarity 94.1%; Pred. No. 8.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 TTACTTCATAGTCTTTG 19
||| ||||| ||||| |||||
Db 123271 TCACITTCATAGTCTTTG 123255

RESULT 82
US-09-949-002-802/c
; Sequence 802, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 802
; LENGTH: 209632
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(209632)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-002-802

; LOCATION: (1)...(151261)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13242

Query Match      81.1%; Score 15.4; DB 3; Length 151261;
Best Local Similarity 94.1%; Pred. No. 8.2e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 TTACTTCATAGTCTTTG 19
||| ||||| ||||| |||||
Db 63211 TTCCTTCATAGTCTTTG 63227

RESULT 81
US-09-949-002-574/c
; Sequence 574, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 574
; LENGTH: 209631
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(209631)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-002-574

Query Match      81.1%; Score 15.4; DB 3; Length 209631;
Best Local Similarity 94.1%; Pred. No. 8.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 TTACTTCATAGTCTTTG 19
||| ||||| ||||| |||||
Db 123271 TCACITTCATAGTCTTTG 123255

RESULT 82
US-09-949-002-802/c
; Sequence 802, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 802
; LENGTH: 209632
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(209632)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-002-802

```

```
; LOCATION: (1)...(300598)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-11868

Query Match      81.1%; Score 15.4; DB 3; Length 300598;
Best Local Similarity 94.1%; Pred. No. 8.5e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TTTACTTCATAGTCTTT 18
   ||||| ||||| ||||| |||||
Db 199658 TTTACTTCATAGTATT 199674

RESULT 85
US-09-949-016-14588
; Sequence 14588, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14588
; LENGTH: 302604
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(302604)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14588

Query Match      81.1%; Score 15.4; DB 3; Length 302604;
Best Local Similarity 94.1%; Pred. No. 8.5e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TTTACTTCATAGTCTTT 18
   ||||| ||||| ||||| |||||
Db 249664 TTTACTTCATAGTATT 249680

RESULT 86
US-09-949-016-14589
; Sequence 14589, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14589
; LENGTH: 302604
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(302604)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14589

Query Match      81.1%; Score 15.4; DB 3; Length 302604;
Best Local Similarity 94.1%; Pred. No. 8.5e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TTTACTTCATAGTCTTT 18
   ||||| ||||| ||||| |||||
Db 249664 TTTACTTCATAGTATT 249680

RESULT 87
US-09-949-016-17119
; Sequence 17119, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17119
; LENGTH: 308362
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(308362)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17119

Query Match      81.1%; Score 15.4; DB 3; Length 308362;
Best Local Similarity 94.1%; Pred. No. 8.5e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TTTACTTCATAGTCTTT 18
   ||||| ||||| ||||| |||||
Db 249480 TTTACTTCATAGTATT 249496

RESULT 88
US-09-949-016-13498
; Sequence 13498, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 13498
; LENGTH: 343352
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(343352)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13498

Query Match      81.1%; Score 15.4; DB 3; Length 343352;
Best Local Similarity 94.1%; Pred. No. 8.5e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      2 TTTACTTCATAGTCTTT 18
Db      283626 TTTACTTCATAGTCTTT 283642

RESULT 89
US-09-949-016-14543/c
; Sequence 14543, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14543
; LENGTH: 387902
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(387902)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14543

Query Match      81.1%; Score 15.4; DB 3; Length 387902;
Best Local Similarity 94.1%; Pred. No. 8.5e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      2 TTTACTTCATAGTCTTT 18
Db      297282 TTTACTTCATAGTCTTT 297266

RESULT 90
US-09-949-016-16297
; Sequence 16297, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/125,219
; PRIOR FILING DATE: 1999-03-18

; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16297
; LENGTH: 421118
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(421118)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16297

Query Match      81.1%; Score 15.4; DB 3; Length 421118;
Best Local Similarity 94.1%; Pred. No. 8.5e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      3 TTACTTCATAGTCTTTG 19
Db      368167 TTACTTCATAGTCTTTG 368183

RESULT 91
US-09-949-016-12557/c
; Sequence 12557, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12557
; LENGTH: 421883
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(421883)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12557

Query Match      81.1%; Score 15.4; DB 3; Length 421883;
Best Local Similarity 94.1%; Pred. No. 8.5e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      2 TTTACTTCATAGTCTTT 18
Db      297282 TTTACTTCATAGTCTTT 297266

RESULT 92
US-09-531-120-211
; Sequence 211, Application US/09531120
; Patent No. 6972197
; GENERAL INFORMATION:
; APPLICANT: PREUSS, DAPHNE
; APPLICANT: COPENHAVER, GREGORY
; TITLE OF INVENTION: PLANT ARTIFICIAL CHROMOSOME COMPOSITIONS AND METHODS
; FILE REFERENCE: ARCD:309US
; CURRENT APPLICATION NUMBER: US/09/531,120
; CURRENT FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/125,219
; PRIOR FILING DATE: 1999-03-18
```

```
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 211
; LENGTH: 1082144
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-531-120-211

Query Match      81.1%; Score 15.4; DB 4; Length 1082144;
Best Local Similarity 94.1%; Pred. No. 7.7e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY  2 TTTACTTCATAGTCTTT 18
    |||||
Db  429813 TTTATTTCATAGTCTTT 429829

RESULT 93
US-08-916-421B-1/c
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus
; Patent No. 6503729
; TITLE OF INVENTION: jannaschii
; FILE REFERENCE: PB275
; CURRENT APPLICATION NUMBER: US/08/916,421B
; CURRENT FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
```

```
; NAME/KEY: misc feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349491)..(1349491)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1470091)..(1470091)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1569020)..(1569020)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1602912)..(1602912)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1603734)..(1603734)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1637998)..(1637998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1664854)..(1664855)
; OTHER INFORMATION: n equals a, t, c, or g
; US-08-916-421B-1
```

```
Query Match 81.1%; Score 15.4; DB 3; Length 1664976;
Best Local Similarity 94.1%; Pred. No. 6.9e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy 2 TTTACTTCATAGTCTTT 18
Db 1303331 TTTACTTCATAGTCTTT 1303315
```

```
RESULT 94
US-09-692-570-1/c
; Sequence 1, Application US/09692570
; Patent No. 6797466
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus
; Patent No. 6797466
; TITLE OF INVENTION: jannaschii
; FILE REFERENCE: PB275C1
; CURRENT APPLICATION NUMBER: US/09/692,570
; CURRENT FILING DATE: 2003-01-14
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
```

```
; PRIOR APPLICATION NUMBER: US 08/916,421
; PRIOR FILING DATE: 1997-08-22
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
```

```
/ NAME/KEY: misc_feature
/ LOCATION: (234187)..(234187)
/ OTHER INFORMATION: n equals a, t, c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (234220)..(234220)
/ OTHER INFORMATION: n equals a, t, c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (234814)..(234814)
/ OTHER INFORMATION: n equals a, t, c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (309398)..(309398)
/ OTHER INFORMATION: n equals a, t, c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (309418)..(309418)
/ OTHER INFORMATION: n equals a, t, c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (312837)..(312837)
/ OTHER INFORMATION: n equals a, t, c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (312993)..(312993)
/ OTHER INFORMATION: n equals a, t, c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (319226)..(319226)
/ OTHER INFORMATION: n equals a, t, c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (559167)..(559167)
/ OTHER INFORMATION: n equals a, t, c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (559241)..(559241)
/ OTHER INFORMATION: n equals a, t, c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (600992)..(600992)
/ OTHER INFORMATION: n equals a, t, c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (622708)..(622708)
/ OTHER INFORMATION: n equals a, t, c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (657081)..(657081)
/ OTHER INFORMATION: n equals a, t, c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (657203)..(657203)
/ OTHER INFORMATION: n equals a, t, c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (674435)..(674435)
/ OTHER INFORMATION: n equals a, t, c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (682442)..(682442)
/ OTHER INFORMATION: n equals a, t, c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (713652)..(713652)
/ OTHER INFORMATION: n equals a, t, c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (741684)..(741684)
/ OTHER INFORMATION: n equals a, t, c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
```

```
/ LOCATION: (779455)..(779455)
/ OTHER INFORMATION: n equals a, t, c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (779676)..(779676)
/ OTHER INFORMATION: n equals a, t, c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (855539)..(855539)
/ OTHER INFORMATION: n equals a, t, c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (871619)..(871619)
/ OTHER INFORMATION: n equals a, t, c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1084830)..(1084830)
/ OTHER INFORMATION: n equals a, t, c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1096846)..(1096846)
/ OTHER INFORMATION: n equals a, t, c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1119881)..(1119881)
/ OTHER INFORMATION: n equals a, t, c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1130881)..(1130881)
/ OTHER INFORMATION: n equals a, t, c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1310988)..(1310988)
/ OTHER INFORMATION: n equals a, t, c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1313224)..(1313224)
/ OTHER INFORMATION: n equals a, t, c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1349473)..(1349473)
/ OTHER INFORMATION: n equals a, t, c, or g
```

Query Match 81.1%; Score 15.4; DB 3; Length 1664976;  
Best Local Similarity 94.1%; Pred. No. 6.9e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TTACTTCATAGTCTTT 18  
|||||  
Db 1303331 TTACTTCATAGTTTTT 1303315

```
RESULT 95
US-08-989-299-3/c
; Sequence 3, Application US/08989299
; Patent No. 6194556
; GENERAL INFORMATION:
; APPLICANT: Acton, Susan L.
; APPLICANT: Robinson, Keith E.
; TITLE OF INVENTION: ANGIOTENSIN CONVERTING ENZYME HOMOLOG
; TITLE OF INVENTION: AND THERAPEUTIC AND DIAGNOSTIC USES THEREFOR
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109-2170
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```



; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/989,299  
; FILING DATE: 11-DEC-1997  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Arnold E., Beth  
; REGISTRATION NUMBER: 35,430  
; REFERENCE/DOCKET NUMBER: MIA-025.01  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-832-1000  
; TELEFAX: 617-832-7000  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2415 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
US-08-989-299-3

Query Match 78.9%; Score 15; DB 3; Length 2415;  
Best Local Similarity 100.0%; Pred. No. 8.9e+02;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TTTACTTCATAGTCT 16  
|||||  
Db 629 TTTACTTCATAGTCT 615

RESULT 96  
US-09-407-427-3/c  
; Sequence 3, Application US/09407427  
; Patent No. 6610497  
; GENERAL INFORMATION:  
; APPLICANT: Acton, Susan L.  
; APPLICANT: Robison, Keith E.  
; TITLE OF INVENTION: ANGIOTENSIN CONVERTING ENZYME HOMOLOG AND THERAPEUTIC  
; FILE REFERENCE: MNI-132CP2  
; CURRENT APPLICATION NUMBER: US/09/407,427  
; CURRENT FILING DATE: 1999-09-29  
; PRIOR APPLICATION NUMBER: 09/163,648  
; PRIOR FILING DATE: 1998-09-30  
; PRIOR APPLICATION NUMBER: 08/989,299  
; PRIOR FILING DATE: 1997-12-11  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 2415  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-407-427-3

Query Match 78.9%; Score 15; DB 3; Length 2415;  
Best Local Similarity 100.0%; Pred. No. 8.9e+02;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TTTACTTCATAGTCT 16  
|||||  
Db 629 TTTACTTCATAGTCT 615

RESULT 97  
US-09-635-501-3/c  
; Sequence 3, Application US/09635501  
; Patent No. 6884771  
; GENERAL INFORMATION:  
; APPLICANT: Acton, Susan L. et al.  
; TITLE OF INVENTION: ANGIOTENSIN CONVERTING ENZYME HOMOLOG AND THERAPEUTIC  
; FILE REFERENCE: MNI-132CP3  
; CURRENT APPLICATION NUMBER: US/09/635,501

; CURRENT FILING DATE: 2000-08-09  
; PRIOR APPLICATION NUMBER: 09/407,427  
; PRIOR FILING DATE: 1999-09-29  
; PRIOR APPLICATION NUMBER: 09/163,648  
; PRIOR FILING DATE: 1998-09-30  
; PRIOR APPLICATION NUMBER: 08/989,299  
; PRIOR FILING DATE: 1997-12-11  
; NUMBER OF SEQ ID NOS: 107  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 2415  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-635-501-3

Query Match 78.9%; Score 15; DB 3; Length 2415;  
Best Local Similarity 100.0%; Pred. No. 8.9e+02;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TTTACTTCATAGTCT 16  
|||||  
Db 629 TTTACTTCATAGTCT 615

RESULT 98  
US-09-163-648-3/c  
; Sequence 3, Application US/09163648  
; Patent No. 6989363  
; GENERAL INFORMATION:  
; APPLICANT: Acton, Susan L. et al.  
; TITLE OF INVENTION: ANGIOTENSIN CONVERTING ENZYME HOMOLOG AND THERAPEUTIC  
; FILE REFERENCE: MNI-132CP  
; CURRENT APPLICATION NUMBER: US/09/163,648  
; CURRENT FILING DATE: 1998-09-30  
; PRIOR APPLICATION NUMBER: 08/989,299  
; PRIOR FILING DATE: 1997-12-11  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 2415  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-163-648-3

Query Match 78.9%; Score 15; DB 4; Length 2415;  
Best Local Similarity 100.0%; Pred. No. 8.9e+02;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TTTACTTCATAGTCT 16  
|||||  
Db 629 TTTACTTCATAGTCT 615

RESULT 99  
US-10-158-847-137/c  
; Sequence 137, Application US/10158847  
; Patent No. 6592865  
; GENERAL INFORMATION:  
; APPLICANT: Tom Parry et al.  
; TITLE OF INVENTION: Method and Compositions for Modulating ACE-2 Activity  
; FILE REFERENCE: PF557  
; CURRENT APPLICATION NUMBER: US/10/158,847  
; CURRENT FILING DATE: 2002-06-03  
; PRIOR APPLICATION NUMBER: 60/295,004  
; PRIOR FILING DATE: 2001-06-04  
; NUMBER OF SEQ ID NOS: 158  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 137  
; LENGTH: 2920  
; TYPE: DNA  
; ORGANISM: homo sapiens  
; FEATURE:

```
; NAME/KEY: misc feature
; LOCATION: (1707)..(1707)
; OTHER INFORMATION: n equals any amino acid
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2702)..(2702)
; OTHER INFORMATION: n equals any amino acid
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2749)..(2749)
; OTHER INFORMATION: n equals any amino acid
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2757)..(2757)
; OTHER INFORMATION: n equals any amino acid
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2788)..(2789)
; OTHER INFORMATION: n equals any amino acid
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2819)..(2819)
; OTHER INFORMATION: n equals any amino acid
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2835)..(2835)
; OTHER INFORMATION: n equals any amino acid
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2856)..(2856)
; OTHER INFORMATION: n equals any amino acid
US-10-158-847-137
```

```
Query Match 78.9%; Score 15; DB 3; Length 2920;
Best Local Similarity 100.0%; Pred. No. 9.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 2 TTACTTCATAGTCT 16
Db 658 TTACTTCATAGTCT 644
```

```
RESULT 100
US-10-158-825-137/c
; Sequence 137, Application US/10158825
; Patent No. 6900033
; GENERAL INFORMATION:
; APPLICANT: Tom Parry et al.
; TITLE OF INVENTION: Method and Compositions for Modulating ACE-2 Activity
; FILE REFERENCE: PF555
; CURRENT APPLICATION NUMBER: US/10/158,825
; CURRENT FILING DATE: 2002-06-03
; PRIOR APPLICATION NUMBER: 60/294,976
; PRIOR FILING DATE: 2001-06-04
; NUMBER OF SEQ ID NOS: 158
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 137
; LENGTH: 2920
; TYPE: DNA
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1707)..(1707)
; OTHER INFORMATION: n equals any amino acid
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2702)..(2702)
; OTHER INFORMATION: n equals any amino acid
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2749)..(2749)
; OTHER INFORMATION: n equals any amino acid
; FEATURE:
```

```
; NAME/KEY: misc feature
; LOCATION: (2757)..(2757)
; OTHER INFORMATION: n equals any amino acid
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2788)..(2789)
; OTHER INFORMATION: n equals any amino acid
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2819)..(2819)
; OTHER INFORMATION: n equals any amino acid
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2835)..(2835)
; OTHER INFORMATION: n equals any amino acid
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2856)..(2856)
; OTHER INFORMATION: n equals any amino acid
US-10-158-825-137
```

```
Query Match 78.9%; Score 15; DB 3; Length 2920;
Best Local Similarity 100.0%; Pred. No. 9.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 2 TTACTTCATAGTCT 16
Db 658 TTACTTCATAGTCT 644
```

```
Search completed: June 10, 2006, 15:37:51
Job time : 81.5917 secs
```

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2006, 15:26:46 ; Search time 727.7 Seconds  
(without alignments) 320.826 Million cell upd

Title: US-10-600-816-32

Perfect score:

Sequence: 1 cttacttcatagtctttg 19

Scoring cable: IDENTITY NUC

scoring curve: IDENTIFY\_NOC  
Gapop 10.0 ; Gapext 1.0

Searched: 18892170 seqs. 6143817638 residues

Total number of hits satisfying chosen parameters: 37784340

Minimum DB seq length: 0

Minimum	DB	req	length:	0
Maximum	DB	req	length:	2000000000

Post-processing: Minimum Match 0%

root processing. Minimum Match 0%  
Maximum Match 100%

Database : Published Applications NA Main:\*

```

1:  /EMC_Celerra_SIDS3/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2:  /EMC_Celerra_SIDS3/ptodata/2/pubpna/US08_PUBCOMB.seq.*
3:  /EMC_Celerra_SIDS3/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
4:  /EMC_Celerra_SIDS3/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
5:  /EMC_Celerra_SIDS3/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
6:  /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
7:  /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
8:  /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
9:  /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10D_PUBCOMB.seq.*
10: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10E_PUBCOMB.seq.*
11: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10F_PUBCOMB.seq.*
12: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10G_PUBCOMB.seq.*
13: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US11A_PUBCOMB.seq.*
14: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US11B_PUBCOMB.seq.*
15: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US11C_PUBCOMB.seq.*
16: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US11D_PUBCOMB.seq.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query		DB	ID	Description
	Score	Match Length			
1	19	100.0	19	8	US-10-600-816-32
C 2	19	100.0	19	14	US-11-083-784-618880
C 3	19	100.0	19	15	US-11-101-244-618880
C 4	19	100.0	42	8	US-10-600-816-30
C 5	19	100.0	497	6	US-10-066-543-1937
C 6	19	100.0	552	6	US-10-066-543-1811
7	19	100.0	620	3	US-09-969-034-2222
8	19	100.0	634	3	US-09-969-034-3393
9	19	100.0	642	7	US-10-125-968-701
C 10	19	100.0	1071	9	US-10-712-615-134
C 11	19	100.0	1212	3	US-09-866-050A-249
C 12	19	100.0	1212	6	US-10-152-661-249
C 13	19	100.0	1328	7	US-10-313-542-223
C 14	19	100.0	1400	13	US-11-060-756-2418
C 15	19	100.0	1400	13	US-11-060-756-6690
C 16	19	100.0	1460	10	US-10-936-636-64
C 17	19	100.0	1460	10	US-10-938-061-64
					Sequence 32, Appl
					Sequence 618880,
					Sequence 618880,
					Sequence 30, Appl
					Sequence 1937, Ap
					Sequence 1811, Ap
					Sequence 2222, Ap
					Sequence 3393, Ap
					Sequence 701, App
					Sequence 134, App
					Sequence 249, App
					Sequence 249, App
					Sequence 223, App
					Sequence 218, A
					Sequence 6690, A
					Sequence 64, App
					Sequence 64, App

C 91	16.4	86.3	3106	10	US-10-750-185-53364	Sequence 53364, A	C 164	15.8	83.2	613	5	US-09-925-065A-45746	Sequence 45746, A
C 92	16.4	86.3	3106	10	US-10-750-623-53364	Sequence 53364, A	C 165	15.8	83.2	613	12	US-10-301-480-146984	Sequence 146984, A
C 93	16	84.2	19	14	US-11-083-784-470217	Sequence 470217, A	C 166	15.8	83.2	613	12	US-10-301-480-760393	Sequence 760393, A
C 94	16	84.2	19	14	US-11-083-784-470217	Sequence 470217, A	C 167	15.8	83.2	614	4	US-09-925-065A-612190	Sequence 612190, A
C 95	16	84.2	19	15	US-11-101-244-470217	Sequence 470217, A	C 168	15.8	83.2	614	5	US-09-925-065A-612190	Sequence 612190, A
C 96	16	84.2	19	15	US-11-101-244-470217	Sequence 470217, A	C 169	15.8	83.2	619	12	US-10-301-480-225030	Sequence 225030, A
C 97	16	84.2	25	13	US-11-060-756-43358	Sequence 43358, A	C 170	15.8	83.2	619	12	US-10-301-480-238439	Sequence 238439, A
C 98	16	84.2	25	13	US-11-060-756-43358	Sequence 43358, A	C 171	15.8	83.2	620	4	US-09-925-065A-128120	Sequence 128120, A
C 99	16	84.2	25	13	US-11-060-756-43361	Sequence 43361, A	C 172	15.8	83.2	620	4	US-09-925-065A-128120	Sequence 128120, A
C 100	16	84.2	366	12	US-10-301-480-242884	Sequence 242884, A	C 173	15.8	83.2	638	4	US-09-925-065A-901546	Sequence 901546, A
C 101	16	84.2	366	12	US-10-301-480-242884	Sequence 242884, A	C 174	15.8	83.2	638	5	US-09-925-065A-901546	Sequence 901546, A
C 102	16	84.2	367	4	US-09-925-065A-148753	Sequence 148753, A	C 175	15.8	83.2	643	6	US-10-027-632-101884	Sequence 101884, A
C 103	16	84.2	367	5	US-09-925-065A-148753	Sequence 148753, A	C 176	15.8	83.2	643	7	US-10-027-632-101884	Sequence 101884, A
C 104	16	84.2	615	4	US-09-925-065A-789256	Sequence 789256, A	C 177	15.8	83.2	645	4	US-09-925-065A-908942	Sequence 908942, A
C 105	16	84.2	615	5	US-09-925-065A-789256	Sequence 789256, A	C 178	15.8	83.2	645	5	US-09-925-065A-908942	Sequence 908942, A
C 106	16	84.2	624	4	US-09-925-065A-549179	Sequence 549179, A	C 179	15.8	83.2	773	12	US-10-301-480-549761	Sequence 549761, A
C 107	16	84.2	624	5	US-09-925-065A-549179	Sequence 549179, A	C 180	15.8	83.2	773	12	US-10-301-480-549761	Sequence 549761, A
C 108	16	84.2	624	12	US-10-301-480-527129	Sequence 527129, A	C 181	15.8	83.2	773	12	US-10-301-480-1163170	Sequence 1163170, A
C 109	16	84.2	624	12	US-10-301-480-527129	Sequence 527129, A	C 182	15.8	83.2	1236	8	US-10-424-599-58252	Sequence 58252, A
C 110	16	84.2	782	10	US-10-750-185-47810	Sequence 47810, A	C 183	15.8	83.2	1258	9	US-10-487-561-20	Sequence 20, Appl
C 111	16	84.2	782	10	US-10-750-185-47810	Sequence 47810, A	C 184	15.8	83.2	1259	3	US-09-764-864-111	Sequence 111, Appl
C 112	16	84.2	782	10	US-10-750-623-47810	Sequence 47810, A	C 185	15.8	83.2	1417	4	US-09-925-065A-74229	Sequence 74229, A
C 113	16	84.2	1120	5	US-09-925-065A-89841	Sequence 89841, A	C 186	15.8	83.2	1417	5	US-09-925-065A-74229	Sequence 74229, A
C 114	16	84.2	1120	5	US-09-925-065A-89841	Sequence 89841, A	C 187	15.8	83.2	1417	5	US-10-301-480-175468	Sequence 175468, A
C 115	16	84.2	1120	12	US-10-301-480-131082	Sequence 131082, A	C 188	15.8	83.2	1417	12	US-10-301-480-788877	Sequence 788877, A
C 116	16	84.2	1120	12	US-10-301-480-804491	Sequence 804491, A	C 189	15.8	83.2	1599	7	US-10-012-697-14565	Sequence 14565, Ap
C 117	16	84.2	1342	4	US-09-925-065A-89456	Sequence 89456, A	C 190	15.8	83.2	1599	10	US-10-779-543-23465	Sequence 23465, A
C 118	16	84.2	1342	5	US-09-925-065A-89456	Sequence 89456, A	C 191	15.8	83.2	1634	8	US-10-437-963-832	Sequence 832, App
C 119	16	84.2	1342	12	US-10-301-480-190697	Sequence 190697, A	C 192	15.8	83.2	1731	10	US-10-450-763-557	Sequence 557, App
C 120	16	84.2	1342	12	US-10-301-480-804106	Sequence 804106, A	C 193	15.8	83.2	1758	7	US-10-295-027-213	Sequence 213, App
C 121	16	84.2	3777	10	US-10-491-472-88	Sequence 88, Appl	C 194	15.8	83.2	2000	8	US-10-260-238-2386	Sequence 2386, Ap
C 122	16	84.2	103391	10	US-10-981-277-54	Sequence 54, Appl	C 195	15.8	83.2	2094	13	US-11-097-143-28058	Sequence 28058, A
C 123	16	84.2	127567	10	US-10-737-082-47	Sequence 47, Appl	C 196	15.8	83.2	2165	13	US-11-097-143-5477	Sequence 5477, Ap
C 124	15.8	83.2	19	14	US-11-083-784-1167118	Sequence 1167118, A	C 197	15.8	83.2	2175	6	US-10-363-829-132	Sequence 1023, Ap
C 125	15.8	83.2	19	15	US-11-101-244-1167118	Sequence 1167118, A	C 198	15.8	83.2	2412	7	US-10-037-270-1023	Sequence 1023, Ap
C 126	15.8	83.2	440	4	US-09-925-065A-358198	Sequence 358198, A	C 199	15.8	83.2	2412	10	US-10-117-722-1023	Sequence 1023, Ap
C 127	15.8	83.2	440	5	US-09-925-065A-358198	Sequence 358198, A	C 200	15.8	83.2	2412	10	US-10-122-851-1023	Sequence 1023, Ap
C 128	15.8	83.2	442	12	US-10-301-480-429467	Sequence 429467, A	C 201	15.8	83.2	2438	9	US-10-357-930-24567	Sequence 24567, A
C 129	15.8	83.2	442	12	US-10-301-480-1042876	Sequence 1042876, A	C 202	15.8	83.2	2438	9	US-10-357-930-24567	Sequence 24567, A
C 130	15.8	83.2	464	8	US-09-925-065A-139697	Sequence 139697, A	C 203	15.8	83.2	2441	13	US-11-097-143-23815	Sequence 23815, A
C 131	15.8	83.2	474	4	US-09-925-065A-781554	Sequence 781554, A	C 204	15.8	83.2	2457	8	US-10-424-599-7103	Sequence 7103, Ap
C 132	15.8	83.2	474	4	US-09-925-065A-842696	Sequence 842696, A	C 205	15.8	83.2	2458	10	US-10-450-763-22230	Sequence 22230, A
C 133	15.8	83.2	474	5	US-09-925-065A-781554	Sequence 781554, A	C 206	15.8	83.2	2935	8	US-10-437-963-48256	Sequence 48256, A
C 134	15.8	83.2	474	5	US-09-925-065A-842696	Sequence 842696, A	C 207	15.8	83.2	3384	8	US-10-437-963-73087	Sequence 73087, A
C 135	15.8	83.2	480	6	US-10-027-632-283411	Sequence 283411, A	C 208	15.8	83.2	10195	3	US-09-764-864-1600	Sequence 1600, Ap
C 136	15.8	83.2	480	6	US-10-027-632-283411	Sequence 283411, A	C 209	15.8	83.2	12462	13	US-11-097-143-5476	Sequence 5476, Ap
C 137	15.8	83.2	480	7	US-10-027-632-283411	Sequence 283411, A	C 210	15.8	83.2	12462	13	US-11-097-143-28057	Sequence 28057, A
C 138	15.8	83.2	480	7	US-10-027-632-283411	Sequence 283411, A	C 211	15.8	83.2	12462	13	US-11-097-143-28057	Sequence 28057, A
C 139	15.8	83.2	500	4	US-09-925-065A-561856	Sequence 561856, A	C 212	15.8	83.2	27393	9	US-10-322-696-133	Sequence 133, App
C 140	15.8	83.2	500	4	US-09-925-065A-561857	Sequence 561857, A	C 213	15.8	83.2	87394	9	US-10-810-788A-6	Sequence 6, Appli
C 141	15.8	83.2	500	5	US-09-925-065A-561857	Sequence 561857, A	C 214	15.8	83.2	117329	10	US-10-987-384-1	Sequence 1, Appli
C 142	15.8	83.2	500	5	US-09-925-065A-561857	Sequence 561857, A	C 215	15.8	83.2	127678	10	US-10-461-862-9	Sequence 9, Appli
C 143	15.8	83.2	573	10	US-10-972-079-76087	Sequence 76087, A	C 216	15.8	83.2	169739	8	US-10-450-826-93	Sequence 93, Appl
C 144	15.8	83.2	578	4	US-09-925-065A-880411	Sequence 880411, A	C 217	15.8	83.2	207835	15	US-11-121-086-39	Sequence 39, Appl
C 145	15.8	83.2	588	5	US-09-925-065A-880411	Sequence 880411, A	C 218	15.8	83.2	207835	15	US-11-121-086-40	Sequence 40, Appl
C 146	15.8	83.2	593	4	US-09-925-065A-632752	Sequence 632752, A	C 219	15.8	83.2	207835	15	US-11-121-086-31	Sequence 31, Appl
C 147	15.8	83.2	593	4	US-09-925-065A-632752	Sequence 632752, A	C 220	15.8	83.2	207835	15	US-10-322-696-31	Sequence 97, Appl
C 148	15.8	83.2	593	5	US-09-925-065A-632752	Sequence 632752, A	C 221	15.8	83.2	207835	6	US-10-388-838-58	Sequence 58, Appl
C 149	15.8	83.2	593	5	US-09-925-065A-632752	Sequence 632752, A	C 222	15.8	83.2	233060	9	US-10-600-816-35	Sequence 35, Appl
C 150	15.8	83.2	593	12	US-10-301-480-335909	Sequence 335909, A	C 223	15.8	83.2	233060	9	US-10-600-816-35	Sequence 35, Appl
C 151	15.8	83.2	593	12	US-10-301-480-335909	Sequence 335909, A	C 224	15.8	83.2	233060	9	US-10-719-900-404459	Sequence 404459, A
C 152	15.8	83.2	596	4	US-09-925-065A-896407	Sequence 896407, A	C 225	15.4	81.1	193	16	US-10-424-599-134281	Sequence 134281, A
C 153	15.8	83.2	596	4	US-09-925-065A-896407	Sequence 896407, A	C 226	15.4	81.1	200	15	US-11-098-686-384	Sequence 384, App
C 154	15.8	83.2	598	12	US-10-301-480-335907	Sequence 335907, A	C 227	15.4	81.1	201	15	US-11-124-368A-11591	Sequence 11591, A
C 155	15.8	83.2	598	12	US-10-301-480-949316	Sequence 949316, A	C 228	15.4	81.1	339	12	US-09-918-995-18700	Sequence 236382, A
C 156	15.8	83.2	600	10	US-10-972-079-10032	Sequence 10032, A	C 229	15.4	81.1	339	12	US-10-301-480-236382	Sequence 849791, A
C 157	15.8	83.2	600	10	US-10-972-079-10033	Sequence 10033, A	C 230	15.4	81.1	351	12	US-10-301-480-849791	Sequence 157970, A
C 158	15.8	83.2	602	4	US-09-925-065A-566918	Sequence 566918, A	C 231	15.4	81.1	354	4	US-09-925-065A-141510	Sequence 141510, A
C 159	15.8	83.2	602	4	US-09-925-065A-566918	Sequence 566918, A	C 232	15.4	81.1	354	5	US-09-925-065A-141510	Sequence 141510, A
C 160	15.8	83.2	602	4	US-09-925-065A-566918	Sequence 566918, A	C 233	15.4	81.1	426	6	US-10-198-846-3285	Sequence 3285, Ap
C 161	15.8	83.2	602	5	US-09-925-065A-566919	Sequence 566919, A	C 234	15.4	81.1	451	3	US-09-876-889-141	Sequence 141, App
C 162	15.8	83.2	602	5	US-09-925-065A-566919	Sequence 566919, A	C 235	15.4	81.1	453	3	US-09-864-761-2102	Sequence 2102, Ap
C 163	15.8	83.2	613	4	US-09-925-065A-45746	Sequence 45746, A	C 236	15.4	81.1	456	3	US-09-918-995-27565	Sequence 27565, A
										459	4	US-09-925-065A-184706	Sequence 184706, A

237	15.4	81.1	459	5	US-09-925-065A-184706	Sequence 184706,	310	15.4	81.1	598	6	US-10-027-632-243582	Sequence 243582,
238	15.4	81.1	461	12	US-10-301-480-274978	Sequence 274978,	311	15.4	81.1	598	7	US-10-027-632-243582	Sequence 243582,
239	15.4	81.1	461	12	US-10-301-480-888387	Sequence 888387,	312	15.4	81.1	599	12	US-10-301-480-33908	Sequence 33908,
240	15.4	81.1	482	3	US-09-918-995-29074	Sequence 29074, A	313	15.4	81.1	599	12	US-10-301-480-949317	Sequence 949317,
241	15.4	81.1	494	5	US-09-925-065A-141505	Sequence 141505,	c 314	15.4	81.1	600	10	US-10-972-079-29176	Sequence 29176, A
242	15.4	81.1	494	5	US-09-925-065A-141505	Sequence 141505,	c 315	15.4	81.1	600	10	US-10-972-079-54266	Sequence 54266, A
243	15.4	81.1	498	12	US-10-301-480-236377	Sequence 236377,	c 316	15.4	81.1	600	10	US-10-972-079-54267	Sequence 54267, A
244	15.4	81.1	498	12	US-10-301-480-849786	Sequence 849786,	c 317	15.4	81.1	600	10	US-10-972-079-54268	Sequence 54268, A
245	15.4	81.1	523	16	US-11-128-061-2442	Sequence 2442, Ap	c 318	15.4	81.1	605	7	US-10-125-968-37	Sequence 27, App1
246	15.4	81.1	523	16	US-11-128-061-6084	Sequence 6084, Ap	c 319	15.4	81.1	608	6	US-10-027-632-193290	Sequence 193290,
247	15.4	81.1	523	16	US-11-128-049-2442	Sequence 2442, Ap	c 320	15.4	81.1	608	6	US-10-027-632-193291	Sequence 193291,
248	15.4	81.1	523	16	US-11-128-049-6084	Sequence 6084, Ap	c 321	15.4	81.1	608	7	US-10-027-632-193290	Sequence 193290,
249	15.4	81.1	525	9	US-10-425-115-43128	Sequence 43128, A	c 322	15.4	81.1	608	7	US-10-027-632-193291	Sequence 193291,
250	15.4	81.1	543	12	US-10-301-480-429464	Sequence 429464,	323	15.4	81.1	611	6	US-10-198-846-13671	Sequence 13671, A
251	15.4	81.1	543	12	US-10-301-480-1042873	Sequence 1042873,	324	15.4	81.1	611	12	US-10-301-480-74819	Sequence 74819, A
252	15.4	81.1	544	4	US-09-925-065A-182736	Sequence 182736,	325	15.4	81.1	611	12	US-10-301-480-74820	Sequence 74820, A
253	15.4	81.1	544	4	US-09-925-065A-182737	Sequence 182737,	326	15.4	81.1	611	12	US-10-301-480-688228	Sequence 688228,
254	15.4	81.1	544	4	US-09-925-065A-182738	Sequence 182738,	327	15.4	81.1	611	12	US-10-301-480-688229	Sequence 688229,
255	15.4	81.1	544	5	US-09-925-065A-182736	Sequence 182736,	c 328	15.4	81.1	613	12	US-10-301-480-424040	Sequence 424040,
256	15.4	81.1	544	5	US-09-925-065A-182737	Sequence 182737,	c 329	15.4	81.1	613	12	US-10-301-480-1037449	Sequence 1037449,
257	15.4	81.1	544	5	US-09-925-065A-182738	Sequence 182738,	c 330	15.4	81.1	616	4	US-09-925-065A-873657	Sequence 873657,
258	15.4	81.1	547	4	US-09-925-065A-231181	Sequence 231181,	c 331	15.4	81.1	616	5	US-09-925-065A-873657	Sequence 873657,
259	15.4	81.1	547	4	US-09-925-065A-231182	Sequence 231182,	332	15.4	81.1	619	4	US-09-925-065A-352227	Sequence 352227,
260	15.4	81.1	547	5	US-09-925-065A-231181	Sequence 231181,	333	15.4	81.1	619	5	US-09-925-065A-352227	Sequence 352227,
261	15.4	81.1	547	5	US-09-925-065A-231182	Sequence 231182,	334	15.4	81.1	624	6	US-10-027-632-234150	Sequence 234150,
262	15.4	81.1	549	4	US-09-925-065A-358194	Sequence 358194,	335	15.4	81.1	624	6	US-10-027-632-234151	Sequence 234151,
263	15.4	81.1	549	5	US-09-925-065A-358194	Sequence 358194,	336	15.4	81.1	624	6	US-10-027-632-234152	Sequence 234152,
264	15.4	81.1	549	12	US-10-301-480-314683	Sequence 314683,	337	15.4	81.1	624	7	US-10-027-632-234150	Sequence 234150,
265	15.4	81.1	549	12	US-10-301-480-314684	Sequence 314684,	338	15.4	81.1	624	7	US-10-027-632-234151	Sequence 234151,
266	15.4	81.1	549	12	US-10-301-480-928092	Sequence 928092,	339	15.4	81.1	624	7	US-10-027-632-234152	Sequence 234152,
267	15.4	81.1	549	12	US-10-301-480-928093	Sequence 928093,	340	15.4	81.1	641	6	US-10-027-632-213247	Sequence 213247,
268	15.4	81.1	550	12	US-10-301-480-273267	Sequence 273267,	341	15.4	81.1	641	7	US-10-027-632-213247	Sequence 213247,
269	15.4	81.1	550	12	US-10-301-480-273268	Sequence 273268,	c 342	15.4	81.1	646	4	US-09-925-065A-530674	Sequence 530674,
270	15.4	81.1	550	12	US-10-301-480-273269	Sequence 273269,	c 343	15.4	81.1	646	5	US-09-925-065A-530674	Sequence 530674,
271	15.4	81.1	550	12	US-10-301-480-886676	Sequence 886676,	344	15.4	81.1	646	12	US-10-301-480-70593	Sequence 70593, A
272	15.4	81.1	550	12	US-10-301-480-886677	Sequence 886677,	345	15.4	81.1	646	12	US-10-301-480-70594	Sequence 70594, A
273	15.4	81.1	550	12	US-10-301-480-886678	Sequence 886678,	346	15.4	81.1	646	12	US-10-301-480-684002	Sequence 684002,
274	15.4	81.1	553	4	US-09-925-065A-141509	Sequence 141509,	347	15.4	81.1	646	12	US-10-301-480-684003	Sequence 684003,
275	15.4	81.1	553	4	US-09-925-065A-331260	Sequence 331260,	348	15.4	81.1	659	4	US-09-925-065A-748225	Sequence 748225,
276	15.4	81.1	553	5	US-09-925-065A-141509	Sequence 141509,	349	15.4	81.1	659	4	US-09-925-065A-748225	Sequence 748225,
c 277	15.4	81.1	558	12	US-10-301-480-236381	Sequence 236381,	350	15.4	81.1	659	6	US-10-027-632-227096	Sequence 227096,
278	15.4	81.1	558	12	US-10-301-480-236381	Sequence 236381,	351	15.4	81.1	659	6	US-10-027-632-227097	Sequence 227097,
279	15.4	81.1	558	12	US-10-301-480-849790	Sequence 849790,	352	15.4	81.1	659	7	US-10-027-632-227096	Sequence 227096,
280	15.4	81.1	562	4	US-09-925-065A-339092	Sequence 339092,	353	15.4	81.1	659	7	US-10-027-632-227097	Sequence 227097,
281	15.4	81.1	562	4	US-09-925-065A-339093	Sequence 339093,	c 354	15.4	81.1	674	4	US-09-925-065A-879545	Sequence 879545,
c 282	15.4	81.1	562	5	US-09-925-065A-339092	Sequence 339092,	c 355	15.4	81.1	674	5	US-09-925-065A-879545	Sequence 879545,
c 283	15.4	81.1	562	5	US-09-925-065A-339093	Sequence 339093,	c 356	15.4	81.1	751	10	US-10-956-157-3668	Sequence 3668, Ap
284	15.4	81.1	573	4	US-09-925-065A-506143	Sequence 506143,	c 357	15.4	81.1	751	10	US-10-956-157-3668	Sequence 8903, Ap
285	15.4	81.1	573	5	US-09-925-065A-506143	Sequence 506143,	c 358	15.4	81.1	811	6	US-10-027-632-167490	Sequence 167490,
286	15.4	81.1	574	12	US-10-301-480-405063	Sequence 405063,	c 359	15.4	81.1	811	7	US-10-027-632-167490	Sequence 167490,
c 287	15.4	81.1	576	12	US-10-301-480-1018472	Sequence 1018472,	c 360	15.4	81.1	845	6	US-10-198-846-7253	Sequence 7253, Ap
c 288	15.4	81.1	576	12	US-10-301-480-412149	Sequence 412149,	361	15.4	81.1	852	8	US-10-437-563-43732	Sequence 43732, A
c 289	15.4	81.1	576	12	US-10-301-480-412150	Sequence 412150,	362	15.4	81.1	873	7	US-10-212-677-234	Sequence 234, App
c 290	15.4	81.1	576	12	US-10-301-480-1025558	Sequence 1025558,	363	15.4	81.1	873	7	US-10-361-811-234	Sequence 234, App
c 291	15.4	81.1	576	12	US-10-301-480-1025559	Sequence 1025559,	364	15.4	81.1	873	7	US-10-369-186-234	Sequence 234, App
c 292	15.4	81.1	580	12	US-10-301-480-80515	Sequence 80515, A	365	15.4	81.1	873	16	US-11-250-759-234	Sequence 234, App
c 293	15.4	81.1	580	12	US-10-301-480-693924	Sequence 693924,	366	15.4	81.1	923	12	US-10-301-480-574262	Sequence 574262,
c 294	15.4	81.1	583	4	US-09-925-065A-321558	Sequence 321558,	367	15.4	81.1	923	12	US-10-301-480-1187671	Sequence 1187671,
c 295	15.4	81.1	583	4	US-09-925-065A-769126	Sequence 769126,	c 368	15.4	81.1	927	8	US-10-282-122A-12470	Sequence 12470, A
c 296	15.4	81.1	583	4	US-09-925-065A-769127	Sequence 769127,	c 369	15.4	81.1	938	16	US-11-079-463-4938	Sequence 4938, Ap
c 297	15.4	81.1	583	5	US-09-925-065A-321558	Sequence 321558	370	15.4	81.1	978	12	US-10-301-480-600611	Sequence 600611,
c 298	15.4	81.1	583	5	US-09-925-065A-769126	Sequence 769126,	371	15.4	81.1	978	12	US-10-301-480-1214020	Sequence 1214020,
c 299	15.4	81.1	583	5	US-09-925-065A-769127	Sequence 769127,	372	15.4	81.1	1053	16	US-11-098-686-8788	Sequence 8788, Ap
c 300	15.4	81.1	583	12	US-10-301-480-396190	Sequence 396190,	c 373	15.4	81.1	1070	8	US-10-425-114-23979	Sequence 23979, A
c 301	15.4	81.1	583	12	US-10-301-480-1009599	Sequence 1009599,	374	15.4	81.1	1190	6	US-10-027-632-116546	Sequence 116546,
302	15.4	81.1	591	4	US-09-925-065A-257224	Sequence 257224,	375	15.4	81.1	1190	6	US-10-027-632-116546	Sequence 116546,
303	15.4	81.1	591	4	US-09-925-065A-257225	Sequence 257225,	c 376	15.4	81.1	1291	10	US-10-750-185-35355	Sequence 35355, A
304	15.4	81.1	591	4	US-09-925-065A-257226	Sequence 257226,	c 377	15.4	81.1	1291	10	US-10-750-185-35355	Sequence 35355, A
305	15.4	81.1	591	4	US-09-925-065A-257227	Sequence 257227,	c 378	15.4	81.1	1333	9	US-10-425-115-48629	Sequence 48629, A
306	15.4	81.1	591	5	US-09-925-065A-257224	Sequence 257224,	c 379	15.4	81.1	1335	4	US-09-925-065A-60635	Sequence 60635, A
307	15.4	81.1	591	5	US-09-925-065A-257225	Sequence 257225,	c 380	15.4	81.1	1335	4	US-09-925-065A-60636	Sequence 60636, A
308	15.4	81.1	591	5	US-09-925-065A-257226	Sequence 257226,	c 381	15.4	81.1	1335	5	US-09-925-065A-60635	Sequence 60635, A
309	15.4	81.1	591	5	US-09-925-065A-257227	Sequence 257227,	c 382	15.4	81.1	1335	5	US-09-925-065A-60636	Sequence 60636, A

c 383	15.4	81.1	1335	12	US-10-301-480-161873	Sequence 161873,	456	15.4	81.1	7848	9	US-10-741-600-246	Sequence 246, App
c 384	15.4	81.1	1335	12	US-10-301-480-161874	Sequence 161874,	457	15.4	81.1	7848	10	US-10-995-561-111	Sequence 111, App
c 385	15.4	81.1	1335	12	US-10-301-480-775282	Sequence 775282,	458	15.4	81.1	7867	10	US-10-098-841-6	Sequence 6, Appli
c 386	15.4	81.1	1335	12	US-10-301-480-775283	Sequence 775283,	459	15.4	81.1	7912	16	US-11-193-771-24	Sequence 24, Appl
c 387	15.4	81.1	1467	16	US-11-079-463-65	Sequence 65, Appl	460	15.4	81.1	7912	16	US-11-193-789-24	Sequence 24, Appl
c 388	15.4	81.1	1562	6	US-10-198-846-11719	Sequence 11719, A	461	15.4	81.1	7912	16	US-11-193-806-24	Sequence 24, Appl
c 389	15.4	81.1	1624	8	US-10-437-963-83034	Sequence 83034, A	462	15.4	81.1	7912	16	US-11-193-857-24	Sequence 24, Appl
c 390	15.4	81.1	1806	7	US-10-369-493-33384	Sequence 33384, A	463	15.4	81.1	7912	16	US-11-193-561-24	Sequence 24, Appl
c 391	15.4	81.1	1825	3	US-09-814-353-20464	Sequence 20464, A	464	15.4	81.1	7935	16	US-10-741-601-74	Sequence 74, Appl
c 392	15.4	81.1	1941	8	US-10-282-1228-16053	Sequence 16053, A	465	15.4	81.1	7935	9	US-10-741-600-240	Sequence 240, App
c 393	15.4	81.1	2028	10	US-10-858-412-79	Sequence 79, Appl	466	15.4	81.1	7935	9	US-10-995-561-113	Sequence 113, App
c 394	15.4	81.1	2163	9	US-10-793-639-382	Sequence 382, Appl	467	15.4	81.1	7959	8	US-10-741-601-81	Sequence 81, Appl
c 395	15.4	81.1	2163	11	US-10-932-182A-78062	Sequence 78062, A	468	15.4	81.1	7959	9	US-10-741-600-249	Sequence 249, App
c 396	15.4	81.1	2197	8	US-10-276-774-839	Sequence 839, App	469	15.4	81.1	7959	10	US-10-995-561-108	Sequence 108, App
c 397	15.4	81.1	2397	7	US-10-104-047-659	Sequence 659, App	470	15.4	81.1	8013	9	US-10-741-601-71	Sequence 71, Appl
c 398	15.4	81.1	2397	16	US-11-072-513-659	Sequence 659, App	471	15.4	81.1	8013	8	US-10-741-600-242	Sequence 242, App
c 399	15.4	81.1	2413	8	US-10-424-599-43603	Sequence 43603, A	472	15.4	81.1	8013	10	US-10-995-561-104	Sequence 104, App
c 400	15.4	81.1	2456	10	US-10-750-185-61032	Sequence 61032, A	473	15.4	81.1	8027	8	US-10-447-161-8	Sequence 8, Appli
c 401	15.4	81.1	2456	10	US-10-750-623-61032	Sequence 61032, A	474	15.4	81.1	8027	8	US-10-734-564-27	Sequence 27, Appl
c 402	15.4	81.1	2494	4	US-09-925-065A-712756	Sequence 712756,	475	15.4	81.1	8027	10	US-10-852-335A-53	Sequence 53, Appl
c 403	15.4	81.1	2494	4	US-09-925-065A-712757	Sequence 712757,	476	15.4	81.1	8027	10	US-10-287-436A-81	Sequence 81, Appl
c 404	15.4	81.1	2494	5	US-09-925-065A-712756	Sequence 712756,	477	15.4	81.1	8044	7	US-10-240-965-121	Sequence 121, App
c 405	15.4	81.1	2494	5	US-09-925-065A-712757	Sequence 712757,	478	15.4	81.1	8044	10	US-10-765-700-135	Sequence 135, App
c 406	15.4	81.1	2532	9	US-10-425-115-106049	Sequence 106049,	479	15.4	81.1	8062	6	US-10-098-841-5	Sequence 5, Appli
c 407	15.4	81.1	2614	4	US-09-925-065A-36691	Sequence 36691, A	480	15.4	81.1	8137	6	US-10-098-841-8	Sequence 8, Appli
c 408	15.4	81.1	2614	5	US-09-925-065A-36691	Sequence 36691, A	481	15.4	81.1	8155	8	US-10-741-601-79	Sequence 79, Appl
c 409	15.4	81.1	2614	12	US-10-301-480-137929	Sequence 137929,	482	15.4	81.1	8155	9	US-10-741-600-247	Sequence 247, App
c 410	15.4	81.1	2614	12	US-10-301-480-751338	Sequence 751338,	483	15.4	81.1	8155	10	US-10-995-561-116	Sequence 116, App
c 411	15.4	81.1	2786	10	US-10-750-185-42848	Sequence 42848, A	484	15.4	81.1	8216	8	US-10-450-763-22270	Sequence 22270, A
c 412	15.4	81.1	2786	10	US-10-750-623-42848	Sequence 42848, A	485	15.4	81.1	8226	8	US-10-741-601-69	Sequence 69, Appl
c 413	15.4	81.1	2817	2	US-08-781-986A-295	Sequence 295, App	486	15.4	81.1	8226	9	US-10-741-600-237	Sequence 237, App
c 414	15.4	81.1	2817	8	US-10-329-624-295	Sequence 295, App	487	15.4	81.1	8226	10	US-10-995-561-107	Sequence 107, App
c 415	15.4	81.1	2873	10	US-10-750-185-24638	Sequence 24638, A	488	15.4	81.1	8230	6	US-10-098-841-7	Sequence 7, Appli
c 416	15.4	81.1	2873	10	US-10-750-623-24638	Sequence 24638, A	489	15.4	81.1	8232	16	US-10-821-234-693	Sequence 693, App
c 417	15.4	81.1	3522	10	US-10-450-763-22266	Sequence 22266, A	490	15.4	81.1	8232	16	US-11-193-771-22	Sequence 22, Appl
c 418	15.4	81.1	3701	7	US-10-388-470-6	Sequence 6, Appli	491	15.4	81.1	8272	16	US-11-193-789-22	Sequence 22, Appl
c 419	15.4	81.1	4013	7	US-10-212-677-267	Sequence 267, App	492	15.4	81.1	8272	16	US-11-193-806-22	Sequence 22, Appl
c 420	15.4	81.1	4013	7	US-10-361-811-267	Sequence 267, App	493	15.4	81.1	8272	16	US-11-193-806-22	Sequence 22, Appl
c 421	15.4	81.1	4013	7	US-10-369-186-267	Sequence 267, App	494	15.4	81.1	8272	16	US-11-193-561-22	Sequence 22, Appl
c 422	15.4	81.1	4013	16	US-11-250-759-267	Sequence 267, App	495	15.4	81.1	8278	8	US-10-741-601-82	Sequence 82, Appl
c 423	15.4	81.1	4259	2	US-08-927-939-30	Sequence 30, Appl	496	15.4	81.1	8278	9	US-10-741-600-250	Sequence 250, App
c 424	15.4	81.1	4259	12	US-10-241-375-30	Sequence 30, Appl	497	15.4	81.1	8278	10	US-10-995-561-106	Sequence 106, App
c 425	15.4	81.1	4385	13	US-11-097-143-8866	Sequence 8866, App	498	15.4	81.1	8332	8	US-10-741-601-73	Sequence 73, Appl
c 426	15.4	81.1	5262	13	US-11-082-454-5	Sequence 5, Appli	499	15.4	81.1	8332	9	US-10-741-600-239	Sequence 239, App
c 427	15.4	81.1	5475	13	US-11-082-454-7	Sequence 7, Appli	500	15.4	81.1	8332	10	US-10-995-561-110	Sequence 110, App
c 428	15.4	81.1	5860	8	US-10-388-470-41	Sequence 41, Appl	501	15.4	81.1	8371	9	US-10-741-601-76	Sequence 76, Appl
c 429	15.4	81.1	6988	8	US-10-236-392-1	Sequence 1, Appli	502	15.4	81.1	8371	9	US-10-741-600-243	Sequence 243, App
c 430	15.4	81.1	7361	8	US-10-236-392-3	Sequence 3, Appli	503	15.4	81.1	8371	10	US-10-995-561-109	Sequence 109, App
c 431	15.4	81.1	7388	16	US-11-193-771-1	Sequence 1, Appli	504	15.4	81.1	8371	16	US-11-193-771-20	Sequence 20, Appl
c 432	15.4	81.1	7388	16	US-11-193-789-1	Sequence 1, Appli	505	15.4	81.1	8374	16	US-11-193-789-20	Sequence 20, Appl
c 433	15.4	81.1	7388	16	US-11-193-806-1	Sequence 1, Appli	506	15.4	81.1	8374	16	US-11-193-806-20	Sequence 20, Appl
c 434	15.4	81.1	7388	16	US-11-193-857-1	Sequence 1, Appli	507	15.4	81.1	8374	16	US-11-193-857-20	Sequence 20, Appl
c 435	15.4	81.1	7388	16	US-11-193-561-1	Sequence 1, Appli	508	15.4	81.1	8374	16	US-11-193-561-20	Sequence 20, Appl
c 436	15.4	81.1	7677	10	US-10-956-157-4995	Sequence 4995, App	509	15.4	81.1	8449	16	US-11-193-771-18	Sequence 18, Appl
c 437	15.4	81.1	7679	10	US-10-831-704-38	Sequence 38, Appl	510	15.4	81.1	8449	16	US-11-193-789-18	Sequence 18, Appl
c 438	15.4	81.1	7680	3	US-09-964-824A-574	Sequence 574, App	511	15.4	81.1	8449	16	US-11-193-806-18	Sequence 18, Appl
c 439	15.4	81.1	7680	6	US-10-171-311-63	Sequence 63, Appl	512	15.4	81.1	8449	16	US-11-193-857-18	Sequence 18, Appl
c 440	15.4	81.1	7680	7	US-10-236-031B-69	Sequence 69, Appl	513	15.4	81.1	8449	16	US-11-193-561-18	Sequence 16, Appl
c 441	15.4	81.1	7680	7	US-10-374-979-75	Sequence 75, Appl	514	15.4	81.1	8647	16	US-11-193-771-16	Sequence 16, Appl
c 442	15.4	81.1	7680	8	US-10-182-336A-75	Sequence 75, Appl	515	15.4	81.1	8647	16	US-11-193-789-16	Sequence 16, Appl
c 443	15.4	81.1	7680	8	US-10-641-643-1289	Sequence 1289, App	516	15.4	81.1	8647	16	US-11-193-806-16	Sequence 16, Appl
c 444	15.4	81.1	7680	8	US-10-717-597-222	Sequence 222, App	517	15.4	81.1	8647	16	US-11-193-857-16	Sequence 16, Appl
c 445	15.4	81.1	7680	9	US-10-788-792-79	Sequence 79, Appl	518	15.4	81.1	8647	16	US-11-193-561-16	Sequence 16, Appl
c 446	15.4	81.1	7680	9	US-10-477-238A-654	Sequence 654, App	519	15.4	81.1	8815	9	US-10-868-577A-62	Sequence 62, Appl
c 447	15.4	81.1	7680	9	US-10-680-287A-654	Sequence 654, App	520	15.4	81.1	8815	9	US-10-868-549-21	Sequence 21, Appl
c 448	15.4	81.1	7680	9	US-10-278-698-88	Sequence 88, Appl	521	15.4	81.1	8815	16	US-11-193-771-14	Sequence 14, Appl
c 449	15.4	81.1	7680	9	US-10-278-698-603	Sequence 603, Appl	522	15.4	81.1	8815	16	US-11-193-789-14	Sequence 14, Appl
c 450	15.4	81.1	7680	10	US-10-843-641A-5877	Sequence 5877, App	523	15.4	81.1	8815	16	US-11-193-806-14	Sequence 14, Appl
c 451	15.4	81.1	7680	10	US-10-477-173-654	Sequence 654, App	524	15.4	81.1	8815	16	US-11-193-857-14	Sequence 14, Appl
c 452	15.4	81.1	7680	10	US-10-852-335A-52	Sequence 52, Appl	525	15.4	81.1	8815	16	US-11-193-857-14	Sequence 14, Appl
c 453	15.4	81.1	7705	8	US-10-447-161-4	Sequence 4, Appli	526	15.4	81.1	14698	13	US-09-764-868-1429	Sequence 1429, App
c 454	15.4	81.1	7795	6	US-10-084-817-2	Sequence 2, Appli	527	15.4	81.1	17815	13	US-11-082-454-54	Sequence 54, Appl
c 455	15.4	81.1	7848	8	US-10-741-601-78	Sequence 78, Appl	528	15.4	81.1	18031	13	US-11-082-454-57	Sequence 57, Appl

529	15.4	81.1	33794	16	US-11-098-686-8737	Sequence 8737, Ap	c 602	15	78.9	3732	6	US-10-175-746-71	Sequence 71, Appl
530	15.4	81.1	87467	8	US-10-741-601-5634	Sequence 5634, Ap	c 603	15	78.9	3732	6	US-10-176-918-71	Sequence 71, Appl
531	15.4	81.1	87467	9	US-10-741-601-17624	Sequence 17624, A	c 604	15	78.9	3732	6	US-10-176-921-71	Sequence 71, Appl
532	15.4	81.1	87672	10	US-10-995-561-13337	Sequence 13237, A	c 605	15	78.9	3732	6	US-10-137-865-71	Sequence 71, Appl
533	15.4	81.1	90361	9	US-10-719-993-6823	Sequence 6823, Ap	c 606	15	78.9	3732	6	US-10-140-474-71	Sequence 71, Appl
534	15.4	81.1	98606	6	US-10-087-192-31	Sequence 31, Appl	c 607	15	78.9	3732	6	US-10-142-431-71	Sequence 71, Appl
535	15.4	81.1	100000	15	US-11-124-368A-2901	Sequence 2901, Ap	c 608	15	78.9	3732	6	US-10-143-114-71	Sequence 71, Appl
536	15.4	81.1	138821	15	US-11-121-086-80	Sequence 80, Appl	c 609	15	78.9	3732	6	US-10-142-419-71	Sequence 71, Appl
537	15.4	81.1	143947	16	US-11-193-771-37	Sequence 37, Appl	c 610	15	78.9	3732	6	US-10-123-262-71	Sequence 71, Appl
538	15.4	81.1	143947	16	US-11-193-789-37	Sequence 37, Appl	c 611	15	78.9	3732	6	US-10-142-423-71	Sequence 71, Appl
539	15.4	81.1	143947	16	US-11-193-806-37	Sequence 37, Appl	c 612	15	78.9	3732	6	US-10-121-050-71	Sequence 71, Appl
540	15.4	81.1	143947	16	US-11-193-857-37	Sequence 37, Appl	c 613	15	78.9	3732	6	US-10-141-755-71	Sequence 71, Appl
541	15.4	81.1	143947	16	US-11-193-561-37	Sequence 37, Appl	c 614	15	78.9	3732	6	US-10-143-032-71	Sequence 71, Appl
542	15.4	81.1	160170	15	US-11-121-086-32	Sequence 32, Appl	c 615	15	78.9	3732	6	US-10-123-108-71	Sequence 71, Appl
543	15.4	81.1	189539	15	US-11-121-086-16	Sequence 16, Appl	c 616	15	78.9	3732	6	US-10-123-236-71	Sequence 71, Appl
544	15.4	81.1	200000	8	US-10-672-764A-30	Sequence 30, Appl	c 617	15	78.9	3732	6	US-10-123-261-71	Sequence 71, Appl
545	15.4	81.1	200000	10	US-10-984-723-1	Sequence 1, Appl	c 618	15	78.9	3732	6	US-10-140-928-71	Sequence 71, Appl
546	15.4	81.1	32938	8	US-10-322-696-79	Sequence 79, Appl	c 619	15	78.9	3732	6	US-10-140-928-71	Sequence 71, Appl
547	15.4	81.1	1082144	15	US-11-117-187-211	Sequence 211, App	c 620	15	78.9	3732	6	US-10-121-045-71	Sequence 71, Appl
548	15	78.9	19	14	US-11-083-784-470128	Sequence 470128, A	c 621	15	78.9	3732	6	US-10-123-292-71	Sequence 71, Appl
549	15	78.9	19	14	US-11-083-784-470228	Sequence 470228, A	c 622	15	78.9	3732	6	US-10-123-903-71	Sequence 71, Appl
550	15	78.9	19	15	US-11-101-244-470128	Sequence 470128, A	c 623	15	78.9	3732	6	US-10-124-819-71	Sequence 71, Appl
551	15	78.9	19	15	US-11-101-244-470228	Sequence 470228, A	c 624	15	78.9	3732	6	US-10-124-819-71	Sequence 71, Appl
552	15	78.9	25	13	US-11-060-756-43360	Sequence 43360, A	c 625	15	78.9	3732	6	US-10-140-925-71	Sequence 71, Appl
553	15	78.9	158	3	US-09-864-761-27572	Sequence 27572, A	c 626	15	78.9	3732	6	US-10-160-498-71	Sequence 71, Appl
554	15	78.9	279	9	US-10-425-115-88166	Sequence 88166, A	c 627	15	78.9	3732	6	US-10-124-824-71	Sequence 71, Appl
555	15	78.9	476	3	US-09-864-761-10935	Sequence 10935, A	c 628	15	78.9	3732	6	US-10-127-825A-71	Sequence 71, Appl
556	15	78.9	487	3	US-09-770-961-748	Sequence 748, App	c 629	15	78.9	3732	6	US-10-127-829A-71	Sequence 71, Appl
557	15	78.9	591	4	US-09-925-065A-556159	Sequence 556159, A	c 630	15	78.9	3732	6	US-10-127-835A-71	Sequence 71, Appl
558	15	78.9	591	5	US-09-925-065A-556159	Sequence 556159, A	c 631	15	78.9	3732	6	US-10-127-839A-71	Sequence 71, Appl
559	15	78.9	599	7	US-10-029-386-102	Sequence 102, App	c 632	15	78.9	3732	6	US-10-127-901A-71	Sequence 71, Appl
560	15	78.9	623	4	US-09-925-065A-867459	Sequence 867459, A	c 633	15	78.9	3732	6	US-10-128-693A-71	Sequence 71, Appl
561	15	78.9	623	5	US-09-925-065A-867459	Sequence 867459, A	c 634	15	78.9	3732	6	US-10-131-813A-71	Sequence 71, Appl
562	15	78.9	635	12	US-10-301-480-53832	Sequence 53832, A	c 635	15	78.9	3732	6	US-10-131-818A-71	Sequence 71, Appl
563	15	78.9	635	12	US-10-301-480-53833	Sequence 53833, A	c 636	15	78.9	3732	6	US-10-131-823A-71	Sequence 71, Appl
564	15	78.9	635	12	US-10-301-480-667241	Sequence 667241, A	c 637	15	78.9	3732	6	US-10-131-824A-71	Sequence 71, Appl
565	15	78.9	635	12	US-10-301-480-667242	Sequence 667242, A	c 638	15	78.9	3732	6	US-10-131-830A-71	Sequence 71, Appl
566	15	78.9	795	16	US-11-079-463-2905	Sequence 2905, Ap	c 639	15	78.9	3732	6	US-10-131-837A-71	Sequence 71, Appl
567	15	78.9	991	12	US-10-301-480-580216	Sequence 580216, A	c 640	15	78.9	3732	6	US-10-137-872A-71	Sequence 71, Appl
568	15	78.9	991	12	US-10-301-480-1193625	Sequence 1193625, A	c 641	15	78.9	3732	6	US-10-147-500-71	Sequence 71, Appl
569	15	78.9	993	12	US-10-301-480-572953	Sequence 572953, A	c 642	15	78.9	3732	6	US-10-147-502-71	Sequence 71, Appl
570	15	78.9	993	12	US-10-301-480-1186362	Sequence 1186362, A	c 643	15	78.9	3732	6	US-10-147-515-71	Sequence 71, Appl
571	15	78.9	1155	3	US-09-938-842A-1650	Sequence 1650, Ap	c 644	15	78.9	3732	6	US-10-147-517-71	Sequence 71, Appl
572	15	78.9	1155	3	US-09-938-842A-1650	Sequence 1650, Ap	c 645	15	78.9	3732	6	US-10-147-526-71	Sequence 71, Appl
573	15	78.9	1183	6	US-10-138-846-10841	Sequence 10841, Ap	c 646	15	78.9	3732	6	US-10-147-527-71	Sequence 71, Appl
574	15	78.9	1760	8	US-10-424-599-78779	Sequence 78779, A	c 647	15	78.9	3732	6	US-10-121-041-71	Sequence 71, Appl
575	15	78.9	2415	13	US-11-059-218-3	Sequence 3, Appl	c 648	15	78.9	3732	6	US-10-121-043-71	Sequence 71, Appl
576	15	78.9	2911	3	US-09-969-384-12	Sequence 12, Appl	c 649	15	78.9	3732	6	US-10-121-047-71	Sequence 71, Appl
577	15	78.9	2920	3	US-09-969-384-2	Sequence 2, Appl	c 650	15	78.9	3732	6	US-10-123-215-71	Sequence 71, Appl
578	15	78.9	2920	6	US-10-158-847-137	Sequence 137, App	c 651	15	78.9	3732	6	US-10-123-902-71	Sequence 71, Appl
579	15	78.9	2920	7	US-10-158-825-137	Sequence 137, App	c 652	15	78.9	3732	6	US-10-123-908-71	Sequence 71, Appl
580	15	78.9	2920	8	US-10-158-825-137	Sequence 137, App	c 653	15	78.9	3732	6	US-10-123-909-71	Sequence 71, Appl
581	15	78.9	3324	6	US-10-116-802-172	Sequence 172, App	c 654	15	78.9	3732	6	US-10-123-910-71	Sequence 71, Appl
582	15	78.9	3325	6	US-10-114-893-85	Sequence 85, Appl	c 655	15	78.9	3732	6	US-10-124-813-71	Sequence 71, Appl
583	15	78.9	3334	3	US-09-978-385-1	Sequence 1, Appl	c 656	15	78.9	3732	6	US-10-124-817-71	Sequence 71, Appl
584	15	78.9	3396	3	US-09-999-781-1	Sequence 1, Appl	c 657	15	78.9	3732	6	US-10-125-922-71	Sequence 71, Appl
585	15	78.9	3396	6	US-10-158-847-141	Sequence 141, App	c 658	15	78.9	3732	6	US-10-125-924-71	Sequence 71, Appl
586	15	78.9	3396	7	US-10-158-825-141	Sequence 141, App	c 659	15	78.9	3732	6	US-10-140-860-71	Sequence 71, Appl
587	15	78.9	3396	10	US-10-158-825-141	Sequence 141, App	c 660	15	78.9	3732	6	US-10-142-417-71	Sequence 71, Appl
588	15	78.9	3396	13	US-11-059-218-1	Sequence 1, Appl	c 661	15	78.9	3732	6	US-10-147-519-71	Sequence 71, Appl
589	15	78.9	3405	6	US-10-005-956-569	Sequence 569, App	c 662	15	78.9	3732	6	US-10-157-872-71	Sequence 71, Appl
590	15	78.9	3405	6	US-10-005-956-842	Sequence 842, App	c 663	15	78.9	3732	6	US-10-152-395-71	Sequence 71, Appl
591	15	78.9	3405	10	US-10-356-157-2371	Sequence 2371, Ap	c 664	15	78.9	3732	6	US-10-125-926A-71	Sequence 71, Appl
592	15	78.9	3405	10	US-10-756-149-2496	Sequence 2496, Ap	c 665	15	78.9	3732	6	US-10-125-930A-71	Sequence 71, Appl
593	15	78.9	3405	10	US-10-798-923-4	Sequence 4, Appl	c 666	15	78.9	3732	6	US-10-127-831A-71	Sequence 71, Appl
594	15	78.9	3405	10	US-10-518-599-1	Sequence 1, Appl	c 667	15	78.9	3732	6	US-10-127-837A-71	Sequence 71, Appl
595	15	78.9	3405	16	US-11-091-883-51	Sequence 51, Appl	c 668	15	78.9	3732	6	US-10-127-838B-71	Sequence 71, Appl
596	15	78.9	3474	10	US-10-204-921-37	Sequence 37, Appl	c 669	15	78.9	3732	6	US-10-127-842A-71	Sequence 71, Appl
597	15	78.9	3732	6	US-10-028-072-71	Sequence 71, Appl	c 670	15	78.9	3732	6	US-10-127-843A-71	Sequence 71, Appl
598	15	78.9	3732	6	US-10-140-808-71	Sequence 71, Appl	c 671	15	78.9	3732	6	US-10-127-845A-71	Sequence 71, Appl
599	15	78.9	3732	6	US-10-121-049-71	Sequence 71, Appl	c 672	15	78.9	3732	6	US-10-127-846A-71	Sequence 71, Appl
600	15	78.9	3732	6	US-10-123-904-71	Sequence 71, Appl	c 673	15	78.9	3732	6	US-10-127-848A-71	Sequence 71, Appl
601	15	78.9	3732	6	US-10-140-470-71	Sequence 71, Appl	c 674	15	78.9	3732	6	US-10-127-849A-71	Sequence 71, Appl







```
c 967      15 78.9 3732 7 US-10-137-867-71 Sequence 71, Appl
c 968      15 78.9 3732 7 US-10-140-473-71 Sequence 71, Appl
c 969      15 78.9 3732 7 US-10-140-806-71 Sequence 71, Appl
c 970      15 78.9 3732 7 US-10-140-810-71 Sequence 71, Appl
c 971      15 78.9 3732 7 US-10-140-863-71 Sequence 71, Appl
c 972      15 78.9 3732 7 US-10-141-699-71 Sequence 71, Appl
c 973      15 78.9 3732 7 US-10-141-703-71 Sequence 71, Appl
c 974      15 78.9 3732 7 US-10-141-706-71 Sequence 71, Appl
c 975      15 78.9 3732 7 US-10-141-757-71 Sequence 71, Appl
c 976      15 78.9 3732 7 US-10-141-762-71 Sequence 71, Appl
c 977      15 78.9 3732 7 US-10-142-428-71 Sequence 71, Appl
c 978      15 78.9 3732 7 US-10-142-429-71 Sequence 71, Appl
c 979      15 78.9 3732 7 US-10-142-884-71 Sequence 71, Appl
c 980      15 78.9 3732 7 US-10-143-027-71 Sequence 71, Appl
c 981      15 78.9 3732 7 US-10-143-115-71 Sequence 71, Appl
c 982      15 78.9 3732 7 US-10-144-956-71 Sequence 71, Appl
c 983      15 78.9 3732 7 US-10-144-958-71 Sequence 71, Appl
c 984      15 78.9 3732 7 US-10-145-632-71 Sequence 71, Appl
c 985      15 78.9 3732 7 US-10-145-749-71 Sequence 71, Appl
c 986      15 78.9 3732 7 US-10-145-753-71 Sequence 71, Appl
c 987      15 78.9 3732 7 US-10-145-871-71 Sequence 71, Appl
c 988      15 78.9 3732 7 US-10-145-878-71 Sequence 71, Appl
c 989      15 78.9 3732 7 US-10-146-794-71 Sequence 71, Appl
c 990      15 78.9 3732 7 US-10-147-489-71 Sequence 71, Appl
c 991      15 78.9 3732 7 US-10-147-507-71 Sequence 71, Appl
c 992      15 78.9 3732 7 US-10-147-535-71 Sequence 71, Appl
c 993      15 78.9 3732 7 US-10-147-537-71 Sequence 71, Appl
c 994      15 78.9 3732 7 US-10-152-376-71 Sequence 71, Appl
c 995      15 78.9 3732 7 US-10-152-381-71 Sequence 71, Appl
c 996      15 78.9 3732 7 US-10-152-400-71 Sequence 71, Appl
c 997      15 78.9 3732 7 US-10-153-585-71 Sequence 71, Appl
c 998      15 78.9 3732 7 US-10-157-780-71 Sequence 71, Appl
c 999      15 78.9 3732 7 US-10-157-800-71 Sequence 71, Appl
c1000     15 78.9 3732 7 US-10-157-801-71 Sequence 71, Appl
```

## ALIGNMENTS

```
RESULT 1
US-10-600-816-32
; Sequence 32, Application US/10600816
; Publication No. US20040121362A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: IDENTIFICATION AND MODULATION OF A G-PROTEIN COUPLED RECEPTOR
; TITLE OF INVENTION: (GPCR), RA13, ASSOCIATED WITH CHRONIC OBSTRUCTIVE PULMONARY
; TITLE OF INVENTION: DISEASE (COPD) AND NF-KB AND E-SELECTIN REGULATION
; FILE REFERENCE: D0251 NP
; CURRENT APPLICATION NUMBER: US/10/600,816
; CURRENT FILING DATE: 2003-06-20
; PRIOR APPLICATION NUMBER: U.S. 60/390,850
; PRIOR FILING DATE: 2002-06-20
; PRIOR APPLICATION NUMBER: U.S.60/407,006
; PRIOR FILING DATE: 2002-08-29
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 32
; LENGTH: 19
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-600-816-32
```

```
Query Match      100.0%; Score 19; DB 8; Length 19;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 CTTTACTTCATAGCTTTG 19
      |||||
Db      1 CTTTACTTCATAGCTTTG 19
```

RESULT 2

```
US-11-083-784-618880/c
; Sequence 618880, Application US/11083784
; Publication No. US20050245475A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 134990US
; CURRENT APPLICATION NUMBER: US/11/083,784
; CURRENT FILING DATE: 2005-03-18
; PRIOR APPLICATION NUMBER: US/10/714,333
; PRIOR FILING DATE: 2003-11-14
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 618880
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-083-784-618880
```

```
Query Match      100.0%; Score 19; DB 14; Length 19;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 CTTTACTTCATAGCTTTG 19
      |||||
Db      19 CTTTACTTCATAGCTTTG 1
```

## RESULT 3

```
US-11-101-244-618880/c
; Sequence 618880, Application US/1101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 134990US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 618880
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-618880
```

```
Query Match      100.0%; Score 19; DB 15; Length 19;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 CTTTACTTCATAGCTTTG 19
      |||||
Db      19 CTTTACTTCATAGCTTTG 1
```

```
RESULT 4
US-10-600-816-30/c
; Sequence 30, Application US/10600816
; Publication No. US20040121362A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: IDENTIFICATION AND MODULATION OF A G-PROTEIN COUPLED RECEPTOR
; TITLE OF INVENTION: (GPCR), RA13, ASSOCIATED WITH CHRONIC OBSTRUCTIVE PULMONARY
; TITLE OF INVENTION: DISEASE (COPD) AND NF-KB AND E-SELECTIN REGULATION
; FILE REFERENCE: D0251 NP
; CURRENT APPLICATION NUMBER: US/10/600,816
; CURRENT FILING DATE: 2003-06-20
; PRIOR APPLICATION NUMBER: U.S. 60/390,850
; PRIOR FILING DATE: 2002-06-20
; PRIOR APPLICATION NUMBER: U.S. 60/407,006
; PRIOR FILING DATE: 2002-08-29
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 30
; LENGTH: 42
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-600-816-30

Query Match      100.0%; Score 19; DB 8; Length 42;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19
Db 42 CTTTACTTCATAGTCTTTG 24

RESULT 5
US-10-066-543-1937/c
; Sequence 1937, Application US/10066543
; Publication No. US20030087818A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Xu, Jiangchun
; APPLICANT: Indirias, Carol Yoseph
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Carter, Darrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Smith, Carole L.
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.563
; CURRENT APPLICATION NUMBER: US/10/066,543
; CURRENT FILING DATE: 2002-01-31
; NUMBER OF SEQ ID NOS: 3417
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1937
; LENGTH: 497
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-066-543-1937

Query Match      100.0%; Score 19; DB 6; Length 497;
Best Local Similarity 100.0%; Pred. No. 88;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19
Db 385 CTTTACTTCATAGTCTTTG 367

RESULT 6
US-10-066-543-1811
; Sequence 1811, Application US/10066543
; Publication No. US20030087818A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Indirias, Carol Yoseph
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Carter, Darrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Smith, Carole L.
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.563
; CURRENT APPLICATION NUMBER: US/10/066,543
; CURRENT FILING DATE: 2002-01-31
; NUMBER OF SEQ ID NOS: 3417
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1811
; LENGTH: 552
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-066-543-1811

Query Match      100.0%; Score 19; DB 6; Length 552;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19
Db 113 CTTTACTTCATAGTCTTTG 131

RESULT 7
US-09-969-034-2222
; Sequence 2222, Application US/09969034
; Publication No. US20040110668A1
; GENERAL INFORMATION:
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Astle, Jon H.
; APPLICANT: Carroll, Eddie III
; APPLICANT: Catino, Theodore J.
; APPLICANT: Dwivedi, Poorima
; APPLICANT: Molino, Gary A.
; APPLICANT: Thiagalingam, Arunthathi
; APPLICANT: Lewis, Marcia E.
; TITLE OF INVENTION: Nucleic Acid Sequences Differentially
; TITLE OF INVENTION: Expressed in Cancer Tissue
; FILE REFERENCE: 1657/1032
; CURRENT APPLICATION NUMBER: US/09/969,034
; CURRENT FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: 60/237,271
; PRIOR FILING DATE: 2000-02-10
; NUMBER OF SEQ ID NOS: 4494
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2222
; LENGTH: 620
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 491, 500, 517, 530, 545, 547, 553, 558, 570, 599, 620
; OTHER INFORMATION: n = A,T,C or G
US-09-969-034-2222

Query Match      100.0%; Score 19; DB 3; Length 620;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19
```

```
Db      113 CTTTACTTCATAGTCTTTG 131
|||||
; LOCATION: 72, 76, 77, 99, 101, 423, 445, 468, 489, 557, 566, 633
; OTHER INFORMATION: n = A,T,C or G
US-10-125-968-701

Query Match      100.0%; Score 19; DB 7; Length 642;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 CTTTACTTCATAGTCTTTG 19
|||||
Db      143 CTTTACTTCATAGTCTTTG 161
|||||

RESULT 10
US-10-712-615-134/c
; Sequence 134, Application US/10712615
; Publication No. US20040214317A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: A NOVEL HUMAN G-PROTEIN COUPLED RECEPTOR, HGPRMY8, EXPRESSED
; FILE REFERENCE: D0047A-CIP
; CURRENT APPLICATION NUMBER: US/10/712,615
; CURRENT FILING DATE: 2003-11-13
; PRIOR APPLICATION NUMBER: U.S. 09/992,238
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: U.S. 60/248,285
; PRIOR FILING DATE: 2000-11-14
; PRIOR APPLICATION NUMBER: U.S. 60/268,581
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: U.S. 60/308,285
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: U.S. 60/317,166
; PRIOR FILING DATE: 2001-09-04
; NUMBER OF SEQ ID NOS: 134
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 134
; LENGTH: 1071
; TYPE: DNA
; ORGANISM: dna:Homo sapiens
US-10-712-615-134

Query Match      100.0%; Score 19; DB 9; Length 1071;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 CTTTACTTCATAGTCTTTG 19
|||||
Db      1059 CTTTACTTCATAGTCTTTG 1041
|||||

RESULT 11
US-09-866-050A-249/c
; Sequence 249, Application US/09866050A
; Publication No. US20030040471A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishanand D.
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; TITLE OF INVENTION: and Methods for Their Use
; FILE REFERENCE: 11000.1011c4U
; CURRENT APPLICATION NUMBER: US/09/866,050A
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 725
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 249
; LENGTH: 1212
; TYPE: DNA
; ORGANISM: Human
```

## US-09-866-050A-249

Query Match 100.0%; Score 19; DB 3; Length 1212;  
Best Local Similarity 100.0%; Pred. No. 99;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCCTTTG 19  
|||||  
Db 1031 CTTTACTTCATAGTCCTTTG 1013

## RESULT 12

US-10-152-661-249/c  
; Sequence 249, Application US/10152661  
; Publication No. US20030022835A1  
; GENERAL INFORMATION:  
; APPLICANT: Watson, James D.  
; APPLICANT: Strachan, Lorna  
; APPLICANT: Sleeman, Matthew  
; APPLICANT: Onrust, Rene  
; APPLICANT: Murison, James G.  
; APPLICANT: Kumble, Krishanand D. Isolated From Skin Cells  
; TITLE OF INVENTION: Compositions and Methods for Their Use  
; FILE REFERENCE: 11000.1011c5  
; CURRENT APPLICATION NUMBER: US/10/152.661  
; CURRENT FILING DATE: 2002-05-20  
; PRIOR APPLICATION NUMBER: 09/866,050  
; PRIOR FILING DATE: 2001-05-24  
; PRIOR APPLICATION NUMBER: 60/221,232  
; PRIOR FILING DATE: 2000-07-25  
; PRIOR APPLICATION NUMBER: 60/206,650  
; PRIOR FILING DATE: 2000-05-24  
; PRIOR APPLICATION NUMBER: 09/312,283  
; PRIOR FILING DATE: 1999-05-14  
; PRIOR APPLICATION NUMBER: PCT/NZ99/00051  
; PRIOR FILING DATE: 1999-04-29  
; PRIOR APPLICATION NUMBER: 09/188,930  
; PRIOR FILING DATE: 1998-11-09  
; PRIOR APPLICATION NUMBER: 09/069,726  
; PRIOR FILING DATE: 1998-04-29  
; NUMBER OF SEQ ID NOS: 725  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 249  
; LENGTH: 1212  
; TYPE: DNA  
; ORGANISM: Human  
US-10-152-661-249

Query Match 100.0%; Score 19; DB 6; Length 1212;  
Best Local Similarity 100.0%; Pred. No. 99;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCCTTTG 19  
|||||  
Db 1031 CTTTACTTCATAGTCCTTTG 1013

## RESULT 13

US-10-313-542-223/c  
; Sequence 223, Application US/10313542  
; Publication No. US20030120057A1  
; GENERAL INFORMATION:  
; APPLICANT: Roopa, Reddy  
; APPLICANT: Guegler, Karl, J.  
; APPLICANT: Au-Young, Janice  
; TITLE OF INVENTION: COMPOSITION FOR DETECTION OF GENES ENCODING MEMBRANE-ASSOCIATED P  
; FILE REFERENCE: PA-0013 US  
; CURRENT APPLICATION NUMBER: US/10/313.542  
; CURRENT FILING DATE: 2002-12-05  
; PRIOR APPLICATION NUMBER: US/09/495,050  
; PRIOR FILING DATE: 2000-01-31  
; PRIOR APPLICATION NUMBER: 60/118,318

; PRIOR FILING DATE: 1999-02-01  
; NUMBER OF SEQ ID NOS: 305  
; SOFTWARE: PERL Program  
; SEQ ID NO 223  
; LENGTH: 1228  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. US20030120057A1 2681738CT1  
US-10-313-542-223

Query Match 100.0%; Score 19; DB 7; Length 1228;  
Best Local Similarity 100.0%; Pred. No. 99;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCCTTTG 19  
|||||  
Db 582 CTTTACTTCATAGTCCTTTG 564

## RESULT 14

US-11-060-756-2418/c  
; Sequence 2418, Application US/11060756  
; Publication No. US20050221354A1  
; GENERAL INFORMATION:  
; APPLICANT: Wyeth  
; APPLICANT: Mounts, William Martin  
; TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug  
; FILE REFERENCE: AM101083 (031896-042000)  
; CURRENT APPLICATION NUMBER: US/11/060,756  
; CURRENT FILING DATE: 2005-02-18  
; NUMBER OF SEQ ID NOS: 303284  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 2418  
; LENGTH: 1400  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-060-756-2418

Query Match 100.0%; Score 19; DB 13; Length 1400;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCCTTTG 19  
|||||  
Db 256 CTTTACTTCATAGTCCTTTG 238

## RESULT 15

US-11-060-756-6690/c  
; Sequence 6690, Application US/11060756  
; Publication No. US20050221354A1  
; GENERAL INFORMATION:  
; APPLICANT: Wyeth  
; APPLICANT: Mounts, William Martin  
; TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug  
; FILE REFERENCE: AM101083 (031896-042000)  
; CURRENT APPLICATION NUMBER: US/11/060,756  
; CURRENT FILING DATE: 2005-02-18  
; NUMBER OF SEQ ID NOS: 303284  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 6690  
; LENGTH: 1400  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-060-756-6690

Query Match 100.0%; Score 19; DB 13; Length 1400;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CTTTACTTCATAGCTTTG 19  
|||||  
Db 256 CTTTACTTCATAGCTTTG 238

## RESULT 16

US-10-936-626-64/c  
; Sequence 64, Application US/10936626  
; Publication No. US20050106644A1  
; GENERAL INFORMATION:  
; APPLICANT: Cairns, Belinda  
; APPLICANT: Chen, Ruihuan  
; APPLICANT: Frantz, Gretchen  
; APPLICANT: Hillan, Kenneth J.  
; APPLICANT: Koepfen, Hartmut  
; APPLICANT: Phillips, Heidi S.  
; APPLICANT: Polakis, Paul  
; APPLICANT: Spencer, Susan D.  
; APPLICANT: Smith, Victoria  
; APPLICANT: Williams, P. Mickey  
; APPLICANT: Wu, Thomas D.  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: Compositions and Methods for the Diagnosis and  
; FILE OF INVENTION: Treatment of Tumor  
; FILE REFERENCE: P5001R1P1  
; CURRENT APPLICATION NUMBER: US/10/936,626  
; PRIOR FILING DATE: 2004-06-21  
; PRIOR APPLICATION NUMBER: US 10/872,991  
; PRIOR FILING DATE: 2004-06-21  
; PRIOR APPLICATION NUMBER: US 10/872,972  
; PRIOR FILING DATE: 2004-06-21  
; PRIOR APPLICATION NUMBER: US 10/241,220  
; PRIOR FILING DATE: 2002-09-11  
; PRIOR APPLICATION NUMBER: US 10/177,488  
; PRIOR FILING DATE: 2002-06-19  
; PRIOR APPLICATION NUMBER: US 60/299,500  
; PRIOR FILING DATE: 2002-06-19  
; PRIOR APPLICATION NUMBER: US 60/301,880  
; PRIOR FILING DATE: 2001-06-20  
; PRIOR APPLICATION NUMBER: US 60/323,268  
; PRIOR FILING DATE: 2001-06-29  
; PRIOR APPLICATION NUMBER: US 60/301,880  
; PRIOR FILING DATE: 2001-06-29  
; PRIOR APPLICATION NUMBER: US 60/323,268  
; PRIOR FILING DATE: 2001-09-18  
; PRIOR APPLICATION NUMBER: US 60/557,116  
; PRIOR FILING DATE: 2004-03-26  
; PRIOR APPLICATION NUMBER: US 60/598,899  
; PRIOR FILING DATE: 2004-08-04  
; NUMBER OF SEQ ID NOS: 154  
; SEQ ID NO 64  
; LENGTH: 1460  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-936-626-64

Query Match 100.0%; Score 19; DB 10; Length 1460;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 CTTTACTTCATAGCTTTG 19  
|||||  
Db 1186 CTTTACTTCATAGCTTTG 1168

## RESULT 17

US-10-938-061-64/c  
; Sequence 64, Application US/10938061  
; Publication No. US20050107595A1  
; GENERAL INFORMATION:  
; APPLICANT: Cairns, Belinda  
; APPLICANT: Chen, Ruihuan  
; APPLICANT: Frantz, Gretchen  
; APPLICANT: Hillan, Kenneth J.  
; APPLICANT: Koepfen, Hartmut  
; APPLICANT: Phillips, Heidi S.

Query Match 100.0%; Score 19; DB 10; Length 1460;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CTTTACTTCATAGCTTTG 19  
|||||  
Db 1186 CTTTACTTCATAGCTTTG 1168

## RESULT 18

US-10-224-289-5/c  
; Sequence 5, Application US/10224289  
; Publication No. US20030207288A1  
; GENERAL INFORMATION:  
; APPLICANT: LEWIN, DAVID A.  
; APPLICANT: STEWART, TIMOTHY A.  
; TITLE OF INVENTION: GPCR-LIKE RETINOIC ACID-INDUCED GENE 1 PROTEIN AND  
; FILE OF INVENTION: NUCLEIC ACID  
; FILE REFERENCE: 9800081-0085  
; CURRENT APPLICATION NUMBER: US/10/224,289  
; CURRENT FILING DATE: 2002-08-20  
; PRIOR APPLICATION NUMBER: 60/313,940  
; PRIOR FILING DATE: 2001-08-20  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 1619  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-224-289-5

Query Match 100.0%; Score 19; DB 7; Length 1619;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

; APPLICANT: Polakis, Paul  
; APPLICANT: Spencer, Susan D.  
; APPLICANT: Smith, Victoria  
; APPLICANT: Williams, P. Mickey  
; APPLICANT: Wu, Thomas D.  
; APPLICANT: Zhang, Zemin  
; APPLICANT: Sakanaka, Chie  
; APPLICANT: Chuntharapal, Anan  
; APPLICANT: Reed Chae J.  
; TITLE OF INVENTION: Compositions and Methods for the Diagnosis and  
; FILE OF INVENTION: Treatment of Tumor  
; FILE REFERENCE: P5001R1P1B  
; CURRENT APPLICATION NUMBER: US/10/938,061  
; CURRENT FILING DATE: 2004-09-10  
; PRIOR APPLICATION NUMBER: US 10/872,991  
; PRIOR FILING DATE: 2004-06-21  
; PRIOR APPLICATION NUMBER: US 10/872,972  
; PRIOR FILING DATE: 2004-06-21  
; PRIOR APPLICATION NUMBER: US 10/241,220  
; PRIOR FILING DATE: 2002-09-11  
; PRIOR APPLICATION NUMBER: US 10/177,488  
; PRIOR FILING DATE: 2002-06-19  
; PRIOR APPLICATION NUMBER: US 60/299,500  
; PRIOR FILING DATE: 2001-06-20  
; PRIOR APPLICATION NUMBER: US 60/301,880  
; PRIOR FILING DATE: 2001-06-29  
; PRIOR APPLICATION NUMBER: US 60/323,268  
; PRIOR FILING DATE: 2001-09-18  
; PRIOR APPLICATION NUMBER: US 60/557,116  
; PRIOR FILING DATE: 2004-03-26  
; PRIOR APPLICATION NUMBER: US 60/598,899  
; PRIOR FILING DATE: 2004-08-04  
; NUMBER OF SEQ ID NOS: 154  
; SEQ ID NO 64  
; LENGTH: 1460  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-938-061-64

Query Match 100.0%; Score 19; DB 10; Length 1460;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CTTTACTTCATAGCTTTG 19  
|||||  
Db 1186 CTTTACTTCATAGCTTTG 1168

## RESULT 18

US-10-224-289-5/c  
; Sequence 5, Application US/10224289  
; Publication No. US20030207288A1  
; GENERAL INFORMATION:  
; APPLICANT: LEWIN, DAVID A.  
; APPLICANT: STEWART, TIMOTHY A.  
; TITLE OF INVENTION: GPCR-LIKE RETINOIC ACID-INDUCED GENE 1 PROTEIN AND  
; FILE OF INVENTION: NUCLEIC ACID  
; FILE REFERENCE: 9800081-0085  
; CURRENT APPLICATION NUMBER: US/10/224,289  
; CURRENT FILING DATE: 2002-08-20  
; PRIOR APPLICATION NUMBER: 60/313,940  
; PRIOR FILING DATE: 2001-08-20  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 1619  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-224-289-5

Query Match 100.0%; Score 19; DB 7; Length 1619;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match 100.0%; Score 19; DB 9; Length 2297;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 CTTTACTTCATAGTCTTTG 19  
|||||  
Db 1178 CTTTACTTCATAGTCTTTG 1160

RESULT 19  
US-10-935-190-43/c  
; Sequence 43, Application US/10935190  
; Publication No. US20050037466A1  
; GENERAL INFORMATION:  
; APPLICANT: INCYTE GENOMICS, INC.  
; APPLICANT: AU-YOUNG, Janice  
; APPLICANT: BANDMAN, Olga  
; APPLICANT: TANG, Y. Tom  
; APPLICANT: YUE, Henry  
; APPLICANT: AZIMZAI, Yalda  
; APPLICANT: BURFORD, Neil  
; APPLICANT: BAUGHN, Mariah R.  
; APPLICANT: LU, Dyung Aina M.  
; APPLICANT: HILLMAN, Jennifer L.  
; APPLICANT: PATTERSON, Chandra  
; APPLICANT: LAL, Preeti  
; TITLE OF INVENTION: RECEPTORS AND ASSOCIATED PROTEINS  
; FILE REFERENCE: PF-0726 PCT  
; CURRENT APPLICATION NUMBER: US/10/935,190  
; CURRENT FILING DATE: 2004-09-08  
; PRIOR APPLICATION NUMBER: US/10/031,904  
; PRIOR FILING DATE: 2002-01-18  
; PRIOR APPLICATION NUMBER: 60/145,232; 60/158,578; 60/165,192  
; PRIOR FILING DATE: 1999-07-21; 1999-10-07; 1999-11-12  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PERL Program  
; SEQ ID NO 43  
; LENGTH: 1619  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: incyte ID No: 2681738CB1  
US-10-935-190-43

Query Match 100.0%; Score 19; DB 9; Length 1619;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 CTTTACTTCATAGTCTTTG 19  
|||||  
Db 1178 CTTTACTTCATAGTCTTTG 1160

RESULT 20  
US-10-775-920-13/c  
; Sequence 13, Application US/10775920  
; Publication No. US2004017544A1  
; GENERAL INFORMATION:  
; APPLICANT: Mergen Ltd  
; TITLE OF INVENTION: POLYNUCLEOTIDE SEQUENCES AND CORRESPONDING ENCODED POLYPEPTIDES  
; TITLE OF INVENTION: OF PARTICULAR SECRETED AND MEMBRANE-BOUND PROTEINS OVEREXPRESSED  
; FILE REFERENCE: IN CERTAIN CANCERS  
; FILE REFERENCE: Mergen - 0010B  
; CURRENT APPLICATION NUMBER: US/10/775,920  
; CURRENT FILING DATE: 2004-02-10  
; PRIOR APPLICATION NUMBER: US 60/447,900  
; PRIOR FILING DATE: 2003-02-13  
; NUMBER OF SEQ ID NOS: 385  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 13  
; LENGTH: 2297  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-775-920-13

Query Match 100.0%; Score 19; DB 9; Length 2297;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 CTTTACTTCATAGTCTTTG 19  
|||||  
Db 1166 CTTTACTTCATAGTCTTTG 1148

RESULT 21  
US-10-224-289-3/c  
; Sequence 3, Application US/10224289  
; Publication No. US20030207288A1  
; GENERAL INFORMATION:  
; APPLICANT: LEWIN, DAVID A.  
; APPLICANT: STEWART, TIMOTHY A.  
; TITLE OF INVENTION: GPCR-LIKE RETINOIC ACID-INDUCED GENE 1 PROTEIN AND  
; FILE REFERENCE: NUCLEIC ACID  
; CURRENT APPLICATION NUMBER: US/10/224,289  
; CURRENT FILING DATE: 2002-08-20  
; PRIOR APPLICATION NUMBER: 60/313,940  
; PRIOR FILING DATE: 2001-08-20  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 2302  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-224-289-3

Query Match 100.0%; Score 19; DB 7; Length 2302;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 CTTTACTTCATAGTCTTTG 19  
|||||  
Db 1158 CTTTACTTCATAGTCTTTG 1140

RESULT 22  
US-10-240-425-405/c  
; Sequence 405, Application US/10240425  
; Publication No. US20040033502A1  
; GENERAL INFORMATION:  
; APPLICANT: Williams, Amanda  
; APPLICANT: Bolland, Joseph F.  
; APPLICANT: Lord, Reginald V.  
; APPLICANT: Alvarez, Chris  
; APPLICANT: Wetzel, Jon C.  
; APPLICANT: Scherf, Uwe  
; APPLICANT: Vockley, Joseph G.  
; TITLE OF INVENTION: Gene Expression Profiles in Esophageal Tissue  
; FILE REFERENCE: 44921-5026  
; CURRENT APPLICATION NUMBER: US/10/240,425  
; CURRENT FILING DATE: 2002-09-30  
; PRIOR APPLICATION NUMBER: PCT/US01/09847  
; PRIOR FILING DATE: 2001-03-28  
; PRIOR APPLICATION NUMBER: US 60/193,446  
; PRIOR FILING DATE: 2000-03-31  
; NUMBER OF SEQ ID NOS: 1588  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 405  
; LENGTH: 2302  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: Genbank Accession No. US20040033502A1 AF095448  
US-10-240-425-405

Query Match 100.0%; Score 19; DB 8; Length 2302;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY      1 CTTTACTTCATAGTCCTTTG 19
      |||||||
Db      1158 CTTTACTTCATAGTCCTTTG 1140

RESULT 23
US-10-775-920-9/c
; Sequence 9, Application US/10775920
; Publication No. US20040175744A1
; GENERAL INFORMATION:
; APPLICANT: Mergen Ltd
; TITLE OF INVENTION: POLYNUCLEOTIDE SEQUENCES AND CORRESPONDING ENCODED POLYPEPTIDES
; TITLE OF INVENTION: OF PARTICULAR SECRETED AND MEMBRANE-BOUND PROTEINS OVEREXPRESSED
; FILE REFERENCE: Merigen - 0010B
; CURRENT APPLICATION NUMBER: US/10/775,920
; CURRENT FILING DATE: 2004-02-10
; PRIOR APPLICATION NUMBER: US 60/447,900
; PRIOR FILING DATE: 2003-02-13
; NUMBER OF SEQ ID NOS: 385
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9
; LENGTH: 2302
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-775-920-9

Query Match      100.0%; Score 19; DB 9; Length 2302;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 CTTTACTTCATAGTCCTTTG 19
      |||||||
Db      1158 CTTTACTTCATAGTCCTTTG 1140

RESULT 24
US-10-510-507-2/c
; Sequence 2, Application US/10510507
; Publication No. US20050282165A1
; GENERAL INFORMATION:
; APPLICANT: Terrett, Jonathan A
; TITLE OF INVENTION: DIAGNOSIS OF CARCINOMA USING RAIG1 POLYPEPTIDES
; FILE REFERENCE: 2543-1-039PCT/US
; CURRENT APPLICATION NUMBER: US/10/510,507
; CURRENT FILING DATE: 2004-10-07
; PRIOR APPLICATION NUMBER: GB0208331.9
; PRIOR FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: GB0221538.2
; PRIOR FILING DATE: 2002-09-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 2302
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-510-507-2

Query Match      100.0%; Score 19; DB 10; Length 2302;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 CTTTACTTCATAGTCCTTTG 19
      |||||||
Db      1158 CTTTACTTCATAGTCCTTTG 1140

RESULT 25
US-10-775-920-12/c
; Sequence 12, Application US/10775920
; Publication No. US20040175744A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Mergen Ltd
; TITLE OF INVENTION: POLYNUCLEOTIDE SEQUENCES AND CORRESPONDING ENCODED POLYPEPTIDES
; TITLE OF INVENTION: OF PARTICULAR SECRETED AND MEMBRANE-BOUND PROTEINS OVEREXPRESSED
; FILE REFERENCE: Merigen - 0010B
; CURRENT APPLICATION NUMBER: US/10/775,920
; CURRENT FILING DATE: 2004-02-10
; PRIOR APPLICATION NUMBER: US 60/447,900
; PRIOR FILING DATE: 2003-02-13
; NUMBER OF SEQ ID NOS: 385
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12
; LENGTH: 2305
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-775-920-12

Query Match      100.0%; Score 19; DB 9; Length 2305;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 CTTTACTTCATAGTCCTTTG 19
      |||||||
Db      1162 CTTTACTTCATAGTCCTTTG 1144

RESULT 26
US-10-176-847-59/c
; Sequence 59, Application US/10176847
; Publication No. US20030068636A1
; GENERAL INFORMATION:
; APPLICANT: Veiby, Petter Ole
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF BREAST
; FILE REFERENCE: MRI-039
; CURRENT APPLICATION NUMBER: US/10/176,847
; CURRENT FILING DATE: 2002-06-21
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 59
; LENGTH: 2316
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-176-847-59

Query Match      100.0%; Score 19; DB 6; Length 2316;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 CTTTACTTCATAGTCCTTTG 19
      |||||||
Db      1164 CTTTACTTCATAGTCCTTTG 1146

RESULT 27
US-11-080-991-59/c
; Sequence 59, Application US/11080991
; Publication No. US20050266437A1
; GENERAL INFORMATION:
; APPLICANT: Veiby, Petter Ole
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF BREAST
; FILE REFERENCE: MRI-039
; CURRENT APPLICATION NUMBER: US/11/080,991
; CURRENT FILING DATE: 2005-03-11
; PRIOR APPLICATION NUMBER: US/10/176,847
; PRIOR FILING DATE: 2002-06-21
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 59
; LENGTH: 2316
```



```
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-080-991-59

Query Match      100.0%; Score 19; DB 15; Length 2316;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 CTTTACTTCATAGCTTTG 19
      |||||
Db      1164 CTTTACTTCATAGCTTTG 1146

RESULT 28
US-10-775-920-11/c
; Sequence 11, Application US/10775920
; Publication No. US2004017544A1
; GENERAL INFORMATION:
; APPLICANT: Merigen Ltd
; TITLE OF INVENTION: POLYNUCLEOTIDE SEQUENCES AND CORRESPONDING ENCODED POLYPEPTIDES
; TITLE OF INVENTION: OF PARTICULAR SECRETED AND MEMBRANE-BOUND PROTEINS OVEREXPRESSED
; TITLE OF INVENTION: IN CERTAIN CANCERS
; FILE REFERENCE: Mergen - 0010B
; CURRENT APPLICATION NUMBER: US/10/775.920
; CURRENT FILING DATE: 2004-02-10
; PRIOR APPLICATION NUMBER: US 60/447,900
; PRIOR FILING DATE: 2003-02-13
; NUMBER OF SEQ ID NOS: 385
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11
; LENGTH: 2446
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-775-920-11

Query Match      100.0%; Score 19; DB 9; Length 2446;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 CTTTACTTCATAGCTTTG 19
      |||||
Db      1312 CTTTACTTCATAGCTTTG 1294

RESULT 29
US-10-225-567A-453/c
; Sequence 453, Application US/10225567A
; Publication No. US20030113798A1
; GENERAL INFORMATION:
; APPLICANT: LifeSpan Biosciences
; APPLICANT: Brown, Joseph P.
; APPLICANT: Burner, Glenna C.
; TITLE OF INVENTION: ANTIGENIC PEPTIDES AND ANTIBODIES FOR G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 1920-4-4
; CURRENT APPLICATION NUMBER: US/10/225.567A
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/257,144
; PRIOR FILING DATE: 2000-12-19
; NUMBER OF SEQ ID NOS: 2292
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 453
; LENGTH: 2456
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-225-567A-453

Query Match      100.0%; Score 19; DB 6; Length 2456;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 CTTTACTTCATAGCTTTG 19
      |||||
```

```
Db      1312 CTTTACTTCATAGCTTTG 1294

RESULT 30
US-10-269-909-63/c
; Sequence 63, Application US/10269909
; Publication No. US20030180747A1
; GENERAL INFORMATION:
; APPLICANT: HRUBAN, RALPH H.
; APPLICANT: ARGANI, PEDRAM
; APPLICANT: IACOBUZIO-DONAHUE, CHRISTINE
; APPLICANT: MAITRA, ANIREBAN
; TITLE OF INVENTION: PANCREATIC CANCER DIAGNOSIS AND THERAPIES
; FILE REFERENCE: 58303(71699)
; CURRENT APPLICATION NUMBER: US/10/269,909
; CURRENT FILING DATE: 2003-10-11
; PRIOR APPLICATION NUMBER: 60/328,609
; PRIOR FILING DATE: 2001-10-11
; PRIOR APPLICATION NUMBER: 60/332,754
; PRIOR FILING DATE: 2001-11-19
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 63
; LENGTH: 2456
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-269-909-63

Query Match      100.0%; Score 19; DB 7; Length 2456;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 CTTTACTTCATAGCTTTG 19
      |||||
Db      1312 CTTTACTTCATAGCTTTG 1294

RESULT 31
US-10-269-909-64/c
; Sequence 64, Application US/10269909
; Publication No. US20030180747A1
; GENERAL INFORMATION:
; APPLICANT: HRUBAN, RALPH H.
; APPLICANT: ARGANI, PEDRAM
; APPLICANT: IACOBUZIO-DONAHUE, CHRISTINE
; APPLICANT: MAITRA, ANIREBAN
; TITLE OF INVENTION: PANCREATIC CANCER DIAGNOSIS AND THERAPIES
; FILE REFERENCE: 58303(71699)
; CURRENT APPLICATION NUMBER: US/10/269,909
; CURRENT FILING DATE: 2003-10-11
; PRIOR APPLICATION NUMBER: 60/328,609
; PRIOR FILING DATE: 2001-10-11
; PRIOR APPLICATION NUMBER: 60/332,754
; PRIOR FILING DATE: 2001-11-19
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 64
; LENGTH: 2456
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-269-909-64

Query Match      100.0%; Score 19; DB 7; Length 2456;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 CTTTACTTCATAGCTTTG 19
      |||||
Db      1312 CTTTACTTCATAGCTTTG 1294

RESULT 32
US-10-295-027-619/c
```

```
; Sequence 619, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevezl, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 619
; LENGTH: 2456
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-295-027-619

Query Match      100.0%; Score 19; DB 7; Length 2456;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 CTTTACTTCATAGTCCTTTG 19
Db      1312 CTTTACTTCATAGTCCTTTG 1294

RESULT 33
US-10-600-816-2/c
; Sequence 2, Application US/10600816
; Publication No. US20040121362A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: IDENTIFICATION AND MODULATION OF A G-PROTEIN COUPLED RECEPTOR
; TITLE OF INVENTION: (GPCR), RAI3, ASSOCIATED WITH CHRONIC OBSTRUCTIVE PULMONARY
; FILE REFERENCE: D0251 NP
; CURRENT APPLICATION NUMBER: US/10/600,816
; CURRENT FILING DATE: 2003-06-20
; PRIOR APPLICATION NUMBER: U.S. 60/390,850
; PRIOR FILING DATE: 2002-06-20
; PRIOR APPLICATION NUMBER: U.S.60/407,006
; PRIOR FILING DATE: 2002-08-29
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 619
; LENGTH: 2456
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-600-816-2/c

Query Match      100.0%; Score 19; DB 8; Length 2456;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 CTTTACTTCATAGTCCTTTG 19
Db      1312 CTTTACTTCATAGTCCTTTG 1294

RESULT 34
US-10-600-816-18/c
; Sequence 18, Application US/10600816
; Publication No. US20040121362A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: IDENTIFICATION AND MODULATION OF A G-PROTEIN COUPLED RECEPTOR
; TITLE OF INVENTION: (GPCR), RAI3, ASSOCIATED WITH CHRONIC OBSTRUCTIVE PULMONARY
; FILE REFERENCE: D0251 NP
; CURRENT APPLICATION NUMBER: US/10/600,816
; CURRENT FILING DATE: 2003-06-20
; PRIOR APPLICATION NUMBER: U.S. 60/390,850
; PRIOR FILING DATE: 2002-06-20
; PRIOR APPLICATION NUMBER: U.S.60/407,006
; PRIOR FILING DATE: 2002-08-29
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 18
; LENGTH: 2456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: RAI3 Polymorphic Allele Summary Sequence.
; NAME/KEY: misc feature
; LOCATION: (112)..(112)
; OTHER INFORMATION: wherein "n" equals either G or A.
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (364)..(364)
; OTHER INFORMATION: wherein "n" equals either C or T.
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (511)..(511)
; OTHER INFORMATION: wherein "n" equals either C or T.
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (523)..(523)
; OTHER INFORMATION: wherein "n" equals either C or T.
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (605)..(605)
; OTHER INFORMATION: wherein "n" equals either A or G.
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (797)..(797)
; OTHER INFORMATION: wherein "n" equals either A or G.
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (111)..(111)
; OTHER INFORMATION: wherein "n" equals either T or C.
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1173)..(1173)
; OTHER INFORMATION: wherein "n" equals either A or G.
US-10-600-816-18
```

```
Query Match      100.0%; Score 19; DB 8; Length 2456;
Best Local Similarity 100.0%; Pred. No. 1.le+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCCTTTG 19
    |||||
Db 1312 CTTTACTTCATAGTCCTTTG 1294

RESULT 35
US-10-775-920-10/c
; Sequence 10, Application US/10775920
; Publication No. US20040175744A1
; GENERAL INFORMATION:
; APPLICANT: Mergen Ltd
; TITLE OF INVENTION: POLYNUCLEOTIDE SEQUENCES AND CORRESPONDING ENCODED POLYPEPTIDES
; TITLE OF INVENTION: OF PARTICULAR SECRETED AND MEMBRANE-BOUND PROTEINS OVEREXPRESSED
; TITLE OF INVENTION: IN CERTAIN CANCERS
; FILE REFERENCE: Mergen - 0010B
; CURRENT APPLICATION NUMBER: US/10/775,920
; CURRENT FILING DATE: 2004-02-10
; PRIOR APPLICATION NUMBER: US 60/447,900
; PRIOR FILING DATE: 2003-02-13
; NUMBER OF SEQ ID NOS: 385
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 10
; LENGTH: 2456
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-775-920-10

Query Match      100.0%; Score 19; DB 9; Length 2456;
Best Local Similarity 100.0%; Pred. No. 1.le+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCCTTTG 19
    |||||
Db 1312 CTTTACTTCATAGTCCTTTG 1294

RESULT 36
US-10-936-626-40/c
; Sequence 40, Application US/10936626
; Publication No. US20050106644A1
; GENERAL INFORMATION:
; APPLICANT: Cairns, Belinda
; APPLICANT: Chen, Ruihuan
; APPLICANT: Frantz, Gretchen
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Koepfen, Hartmut
; APPLICANT: Phillips, Heidi S.
; APPLICANT: Polakis, Paul
; APPLICANT: Spencer, Susan D.
; APPLICANT: Smith, Victoria
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wu, Thomas D.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: Compositions and Methods for the Diagnosis and
; TITLE OF INVENTION: Treatment of Tumor
; FILE REFERENCE: P5001R1P1
; CURRENT APPLICATION NUMBER: US/10/936,626
; CURRENT FILING DATE: 2004-09-08
; PRIOR APPLICATION NUMBER: US 10/872,991
; PRIOR FILING DATE: 2004-06-21
; PRIOR APPLICATION NUMBER: US 10/872,972
; PRIOR FILING DATE: 2004-06-21
; PRIOR APPLICATION NUMBER: US 10/241,220
; PRIOR FILING DATE: 2002-09-11
; PRIOR APPLICATION NUMBER: US 10/177,488
; PRIOR FILING DATE: 2002-06-19
; PRIOR APPLICATION NUMBER: US 60/299,500
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 60/301,880
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/323,268
; PRIOR FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US 60/557,116
; PRIOR FILING DATE: 2004-03-26
; PRIOR APPLICATION NUMBER: US 60/598,899
; PRIOR FILING DATE: 2004-08-04
; NUMBER OF SEQ ID NOS: 154
; SEQ ID NO 40
; LENGTH: 2456
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-936-626-40
```

```
Query Match      100.0%; Score 19; DB 10; Length 2456;
Best Local Similarity 100.0%; Pred. No. 1.le+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCCTTTG 19
    |||||
Db 1312 CTTTACTTCATAGTCCTTTG 1294

RESULT 37
US-10-938-061-40/c
; Sequence 40, Application US/10938061
; Publication No. US20050107595A1
; GENERAL INFORMATION:
; APPLICANT: Cairns, Belinda
; APPLICANT: Chen, Ruihuan
; APPLICANT: Frantz, Gretchen
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Koepfen, Hartmut
; APPLICANT: Phillips, Heidi S.
; APPLICANT: Polakis, Paul
; APPLICANT: Spencer, Susan D.
; APPLICANT: Smith, Victoria
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wu, Thomas D.
; APPLICANT: Zhang, Zemin
; APPLICANT: Sakanaka, Chie
; APPLICANT: Chuntharapai, Anan
; APPLICANT: Reed Chae J.
; TITLE OF INVENTION: Compositions and Methods for the Diagnosis and
; TITLE OF INVENTION: Treatment of Tumor
; FILE REFERENCE: P5001R1P1B
; CURRENT APPLICATION NUMBER: US/10/938,061
; CURRENT FILING DATE: 2004-09-10
; PRIOR APPLICATION NUMBER: US 10/872,991
; PRIOR FILING DATE: 2004-06-21
; PRIOR APPLICATION NUMBER: US 10/872,972
; PRIOR FILING DATE: 2004-06-21
; PRIOR APPLICATION NUMBER: US 10/241,220
; PRIOR FILING DATE: 2002-09-11
; PRIOR APPLICATION NUMBER: US 10/177,488
; PRIOR FILING DATE: 2002-06-19
; PRIOR APPLICATION NUMBER: US 60/299,500
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 60/301,880
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/323,268
; PRIOR FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US 60/557,116
; PRIOR FILING DATE: 2004-03-26
; PRIOR APPLICATION NUMBER: US 60/598,899
; PRIOR FILING DATE: 2004-08-04
; NUMBER OF SEQ ID NOS: 154
; SEQ ID NO 40
; LENGTH: 2456
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-938-061-40
```

Query Match 100.0%; Score 19; DB 10; Length 2456;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCCTTTG 19  
|||||  
Db 1312 CTTTACTTCATAGTCCTTTG 1294

RESULT 38  
US-11-169-041-32/c  
; Sequence 32, Application US/11169041  
; Publication No. US20060019284A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: IDENTIFICATION OF POLYNUCLEOTIDES FOR PREDICTING ACTIVITY OF  
; TITLE OF INVENTION: COMPOUNDS THAT INTERACT WITH AND/OR MODULATE PROTEIN TYROSINE  
; TITLE OF INVENTION: KINASES AND/OR PROTEIN TYROSINE KINASE PATHWAYS IN LUNG CANCER  
; TITLE OF INVENTION: CELLS  
; FILE REFERENCE: 10001 NP  
; CURRENT APPLICATION NUMBER: US/11/169,041  
; CURRENT FILING DATE: 2005-06-28  
; PRIOR APPLICATION NUMBER: 60/584,405  
; PRIOR FILING DATE: 2004-06-30  
; NUMBER OF SEQ ID NOS: 527  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 32  
; LENGTH: 2456  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-169-041-32

Query Match 100.0%; Score 19; DB 16; Length 2456;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCCTTTG 19  
|||||  
Db 1312 CTTTACTTCATAGTCCTTTG 1294

RESULT 39  
US-10-264-049-834/c  
; Sequence 834, Application US/10264049  
; Publication No. US20040005579A1  
; GENERAL INFORMATION:  
; APPLICANT: Birse et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PA133P1  
; CURRENT APPLICATION NUMBER: US/10/264,049  
; CURRENT FILING DATE: 2002-10-04  
; PRIOR APPLICATION NUMBER: PCT/US01/18569  
; PRIOR FILING DATE: 2001-06-07  
; PRIOR APPLICATION NUMBER: US 60/209,467  
; PRIOR FILING DATE: 2000-06-07  
; NUMBER OF SEQ ID NOS: 4360  
; SOFTWARE: PatentIn Ver. 3.1  
; SEQ ID NO 834  
; LENGTH: 2593  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-264-049-834

Query Match 100.0%; Score 19; DB 7; Length 2593;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCCTTTG 19  
|||||  
Db 1424 CTTTACTTCATAGTCCTTTG 1406

RESULT 40  
US-10-198-846-10424/c  
; Sequence 10424, Application US/10198846  
; Publication No. US20030099974A1  
; GENERAL INFORMATION:  
; APPLICANT: Lillie, James  
; APPLICANT: Xu, Yongyao  
; APPLICANT: Wang, Youzhen  
; APPLICANT: Steinmann, Kathleen  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS  
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
; TITLE OF INVENTION: THERAPY OF BREAST CANCER  
; FILE REFERENCE: MRI-049  
; CURRENT APPLICATION NUMBER: US/10/198,846  
; CURRENT FILING DATE: 2002-07-18  
; PRIOR APPLICATION NUMBER: 60/306,220  
; PRIOR FILING DATE: 2001-07-18  
; NUMBER OF SEQ ID NOS: 14084  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10424  
; LENGTH: 4239  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 1, 4238, 4239  
; OTHER INFORMATION: n = A,T,C or G  
US-10-198-846-10424

Query Match 100.0%; Score 19; DB 6; Length 4239;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCCTTTG 19  
|||||  
Db 1720 CTTTACTTCATAGTCCTTTG 1702

RESULT 41  
US-09-925-065A-840554  
; Sequence 840554, Application US/09925065A  
; Publication No. US20040181048A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 840554  
; LENGTH: 523  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-840554

Query Match 92.6%; Score 17.6; DB 4; Length 523;  
Best Local Similarity 94.4%; Pred. No. 4.1e+02;  
Matches 17; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCCTTT 18  
|||||

Db 403 CTTTACTTMTAGTCTTT 420

## RESULT 42

US-09-925-065A-840554  
; Sequence 840554, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925.065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR FILING DATE: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 840554  
; LENGTH: 523  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-840554

Query Match 92.6%; Score 17.6; DB 5; Length 523;  
Best Local Similarity 94.4%; Pred. No. 4.1e+02;  
Matches 17; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTACTTMTAGTCTTT 18  
| | | | | | | | | | | | | | | | | | | |  
Db 403 CTTTACTTMTAGTCTTT 420

## RESULT 43

US-09-925-065A-916369  
; Sequence 916369, Application US/09925065A  
; Publication No. US20040181048A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925.065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR FILING DATE: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 916369  
; LENGTH: 638  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-916369

Query Match 91.6%; Score 17.4; DB 4; Length 638;  
Best Local Similarity 94.7%; Pred. No. 5.2e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19  
| | | | | | | | | | | | | | | | | | | |  
Db 287 CTTTACTTCATAGTCTTTG 305

## RESULT 44

US-09-925-065A-916369  
; Sequence 916369, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925.065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR FILING DATE: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 916369  
; LENGTH: 638  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-916369

Query Match 91.6%; Score 17.4; DB 5; Length 638;  
Best Local Similarity 94.7%; Pred. No. 5.2e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19  
| | | | | | | | | | | | | | | | | | | |  
Db 287 CTTTACTTCATAGTCTTTG 305

## RESULT 45

US-10-301-480-556032/c  
; Sequence 556032, Application US/10301480  
; Publication No. US20060057564A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms  
; in the Human Genome  
; FILE REFERENCE: 108827.137  
; CURRENT APPLICATION NUMBER: US/10/301.480  
; CURRENT FILING DATE: 2002-11-21  
; PRIOR FILING DATE: US 10/215,598  
; PRIOR FILING DATE: 2002-08-09  
; PRIOR APPLICATION NUMBER: US 60/311,695  
; PRIOR FILING DATE: 2001-08-10  
; NUMBER OF SEQ ID NOS: 1226818  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 556032  
; LENGTH: 997  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-10-301-480-556032

Query Match 91.6%; Score 17.4; DB 12; Length 997;  
Best Local Similarity 94.7%; Pred. No. 5.5e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19  
| | | | | | | | | | | | | | | | | | | |

```
Db      417 CTTTACTTCATGTCCTTTG 399

RESULT 46
US-10-301-480-1169441/c
; Sequence 1169441, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1169441
; LENGTH: 997
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-1169441

Query Match      91.6%; Score 17.4; DB 12; Length 997;
Best Local Similarity 94.7%; Pred. No. 5.5e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CTTTACTTCATGTCCTTTG 19
      |||||
Db      417 CTTTACTTCATGTCCTTTG 399

RESULT 47
US-10-437-963-53068/c
; Sequence 53068, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 53068
; LENGTH: 1371
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_55303C.1
US-10-437-963-53068

Query Match      91.6%; Score 17.4; DB 8; Length 1371;
Best Local Similarity 94.7%; Pred. No. 5.7e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CTTTACTTCATGTCCTTTG 19
      |||||
Db      135 CTTTGCTTCATGTCCTTTG 117

RESULT 48
US-09-946-290-7
; Sequence 7, Application US/09946290
; Publication No. US20030211475A1
; GENERAL INFORMATION:
; APPLICANT: Roberts, Christopher J.
; TITLE OF INVENTION: METHODS FOR IDENTIFYING PATHWAY-SPECIFIC REPORTERS AND
; FILE REFERENCE: 9301-040
; CURRENT APPLICATION NUMBER: US/09/946,290
; CURRENT FILING DATE: 2001-09-05
; PRIOR APPLICATION NUMBER: US/09/282,243
; PRIOR FILING DATE: 1999-03-31
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 1725
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1001)..(1522)
US-09-946-290-7

Query Match      91.6%; Score 17.4; DB 3; Length 1725;
Best Local Similarity 94.7%; Pred. No. 5.9e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CTTTACTTCATGTCCTTTG 19
      |||||
Db      386 CTTTACTTCATGTCCTTTG 404

RESULT 49
US-10-505-486-196/c
; Sequence 196, Application US/10505486
; Publication No. US20050118639A1
; GENERAL INFORMATION:
; APPLICANT: Takeda Chemical Industries, Ltd.
; TITLE OF INVENTION: Determination of a ligand
; FILE REFERENCE: P03-0006PCT
; CURRENT APPLICATION NUMBER: US/10/505,486
; CURRENT FILING DATE: 2004-08-20
; PRIOR APPLICATION NUMBER: JP 2002-45728
; PRIOR FILING DATE: 2002-02-22
; PRIOR APPLICATION NUMBER: JP 2002-213949
; PRIOR FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: JP 2002-298237
; PRIOR FILING DATE: 2002-10-11
; NUMBER OF SEQ ID NOS: 233
; SEQ ID NO 196
; LENGTH: 1788
; TYPE: DNA
; ORGANISM: Human
US-10-505-486-196

Query Match      91.6%; Score 17.4; DB 10; Length 1788;
Best Local Similarity 94.7%; Pred. No. 5.9e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CTTTACTTCATGTCCTTTG 19
      |||||
Db      1059 CTTTACTTCATGTCCTTTG 1041

RESULT 50
US-11-116-881A-1152
; Sequence 1152, Application US/11116881A
; Publication No. US20060041949A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Dongmei
; APPLICANT: Nielsen, Mark T.
; TITLE OF INVENTION: Nicotiana Nucleic Acid Molecules and Uses Thereof
; FILE REFERENCE: 07678/141014
; CURRENT APPLICATION NUMBER: US/11/116,881A
```

; CURRENT FILING DATE: 2005-04-27  
; PRIOR APPLICATION NUMBER: 60/665,451  
; FILING DATE: 2005-03-24  
; PRIOR APPLICATION NUMBER: 60/665,097  
; PRIOR FILING DATE: 2005-03-24  
; PRIOR APPLICATION NUMBER: 60/646,764  
; PRIOR FILING DATE: 2005-01-25  
; PRIOR APPLICATION NUMBER: 60/607,357  
; PRIOR FILING DATE: 2004-09-03  
; PRIOR APPLICATION NUMBER: 60/566,235  
; PRIOR FILING DATE: 2004-04-29  
; PRIOR APPLICATION NUMBER: 10/934,944  
; PRIOR FILING DATE: 2004-09-03  
; PRIOR APPLICATION NUMBER: 10/943,507  
; PRIOR FILING DATE: 2004-09-17  
; PRIOR APPLICATION NUMBER: 60/503,989  
; PRIOR FILING DATE: 2003-09-18  
; PRIOR APPLICATION NUMBER: 60/485,368  
; PRIOR FILING DATE: 2003-07-08  
; PRIOR APPLICATION NUMBER: 60/418,933  
; PRIOR FILING DATE: 2002-10-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 2300  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 1152  
; LENGTH: 168  
; TYPE: DNA  
; ORGANISM: Nicotiana tabacum  
US-11-116-881A-1152

Query Match 86.3%; Score 16.4; DB 16; Length 168;  
Best Local Similarity 94.4%; Pred. No. 1.3e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TTACTTCATAGTCTTTG 19  
|||||  
Db 35 TTACTTCATAGTCTTTG 52

RESULT 51  
US-10-301-480-242885  
; Sequence 242885, Application US/10301480  
; Publication No. US20060057564A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms  
; FILE REFERENCE: 108827.137  
; CURRENT APPLICATION NUMBER: US/10/301,480  
; CURRENT FILING DATE: 2002-11-21  
; PRIOR APPLICATION NUMBER: US 10/215,598  
; PRIOR FILING DATE: 2002-08-09  
; PRIOR APPLICATION NUMBER: US 60/311,695  
; PRIOR FILING DATE: 2001-08-10  
; NUMBER OF SEQ ID NOS: 1226818  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 242885  
; LENGTH: 366  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-10-301-480-242885

Query Match 86.3%; Score 16.4; DB 12; Length 366;  
Best Local Similarity 94.4%; Pred. No. 1.4e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TTACTTCATAGTCTTTG 19  
|||||  
Db 44 TTACTTCAAAAGTCTTTG 61

RESULT 52  
US-10-301-480-856294

; Sequence 856294, Application US/10301480  
; Publication No. US20060057564A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms  
; FILE REFERENCE: 108827.137  
; CURRENT APPLICATION NUMBER: US/10/301,480  
; CURRENT FILING DATE: 2002-11-21  
; PRIOR APPLICATION NUMBER: US 10/215,598  
; PRIOR FILING DATE: 2002-08-09  
; PRIOR APPLICATION NUMBER: US 60/311,695  
; PRIOR FILING DATE: 2001-08-10  
; NUMBER OF SEQ ID NOS: 1226818  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 856294  
; LENGTH: 366  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-10-301-480-856294

Query Match 86.3%; Score 16.4; DB 12; Length 366;  
Best Local Similarity 94.4%; Pred. No. 1.4e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TTACTTCATAGTCTTTG 19  
|||||  
Db 44 TTACTTCAAAAGTCTTTG 61

RESULT 53  
US-09-925-065A-148754  
; Sequence 148754, Application US/09925065A  
; Publication No. US20040181048A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 148754  
; LENGTH: 367  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-148754

Query Match 86.3%; Score 16.4; DB 4; Length 367;  
Best Local Similarity 94.4%; Pred. No. 1.4e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TTACTTCATAGTCTTTG 19  
|||||  
Db 44 TTACTTCAAAAGTCTTTG 61

RESULT 54  
US-09-925-065A-148754  
; Sequence 148754, Application US/09925065A  
; Publication No. US2005028172A9  
; GENERAL INFORMATION:

; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; FILE OF INVENTION: Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 148754  
; LENGTH: 367  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-148754

Query Match 86.3%; Score 16.4; DB 5; Length 367;  
Best Local Similarity 94.4%; Pred. No. 1.4e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TTTACTTCATAGCTTTG 19  
Db 44 TTTACTTCAAAGTCTTTG 61

## RESULT 55

US-09-925-065A-500185  
; Sequence 500185, Application US/09925065A  
; Publication No. US20040181049A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; FILE OF INVENTION: Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 500185  
; LENGTH: 572  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-500185

Query Match 86.3%; Score 16.4; DB 4; Length 572;  
Best Local Similarity 94.4%; Pred. No. 1.5e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGCTTTT 18  
Db 230 CTTCACTTCATAGCTTTT 247

## RESULT 56

US-09-925-065A-500185

; Sequence 500185, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; FILE OF INVENTION: Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 500185  
; LENGTH: 572  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-500185

Query Match 86.3%; Score 16.4; DB 5; Length 572;  
Best Local Similarity 94.4%; Pred. No. 1.5e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGCTTTT 18  
Db 230 CTTCACTTCATAGCTTTT 247

## RESULT 57

US-10-972-079-81269  
; Sequence 81269, Application US/10972079  
; Publication No. US20050153317A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: ROSENFELD, David  
; APPLICANT: KERR, Richard  
; APPLICANT: BATES, Stephen  
; APPLICANT: HOLM, Tom  
; TITLE OF INVENTION: METHODS & SYSTEMS FOR INFERRING TRAITS TO BREED & MANAGE NON-BEEF  
; FILE OF INVENTION: LIVESTOCK  
; FILE REFERENCE: MM1110-2  
; CURRENT APPLICATION NUMBER: US/10/972,079  
; CURRENT FILING DATE: 2004-10-22  
; PRIOR APPLICATION NUMBER: US 60/514,333  
; PRIOR FILING DATE: 2003-10-24  
; NUMBER OF SEQ ID NOS: 96631  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 81269  
; LENGTH: 576  
; TYPE: DNA  
; ORGANISM: Chicken 19866894374567\_3  
US-10-972-079-81269

Query Match 86.3%; Score 16.4; DB 10; Length 576;  
Best Local Similarity 94.4%; Pred. No. 1.5e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TTTACTTCATAGCTTTG 19  
Db 528 TTTACTTCTTAGTCTTTG 545

## RESULT 58

US-09-925-065A-951021



; Sequence 951021, Application US/09925065A  
; Publication No. US20040181048A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925.065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 951021  
; LENGTH: 595  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-951021

Query Match 86.3%; Score 16.4; DB 4; Length 595;  
Best Local Similarity 94.4%; Pred. No. 1.5e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCCTTT 18  
||| |||||  
Db 233 CTTCACTTCATAGTCCTTT 250

RESULT 59  
US-09-925-065A-951021  
; Sequence 951021, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925.065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 951021  
; LENGTH: 595  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-951021

Query Match 86.3%; Score 16.4; DB 5; Length 595;  
Best Local Similarity 94.4%; Pred. No. 1.5e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCCTTT 18  
||| |||||  
Db 233 CTTCACTTCATAGTCCTTT 250

RESULT 60  
US-10-425-115-71652  
; Sequence 71652, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53222)B  
; CURRENT APPLICATION NUMBER: US/10/425.115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369326  
; SEQ ID NO 71652  
; LENGTH: 602  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: MRT4577\_165350C.1  
US-10-425-115-71652

Query Match 86.3%; Score 16.4; DB 9; Length 602;  
Best Local Similarity 94.4%; Pred. No. 1.5e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCCTTT 18  
||| |||||  
Db 363 CTTCACTTCATAGTCCTTT 380

RESULT 61  
US-09-925-065A-688576/c  
; Sequence 688576, Application US/09925065A  
; Publication No. US20040181048A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925.065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 688576  
; LENGTH: 680  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-688576

Query Match 86.3%; Score 16.4; DB 4; Length 680;  
Best Local Similarity 94.4%; Pred. No. 1.6e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCCTTT 18  
||| |||||  
Db 476 CTTTCTTCATAGTCCTTT 459

RESULT 62  
US-09-925-065A-688577/c

```
; Sequence 688577, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR FILING DATE: 2000-10-24
; PRIOR FILING DATE: 2000-11-20
; PRIOR FILING DATE: 2000-11-30
; PRIOR FILING DATE: 2001-01-16
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 688577
; LENGTH: 680
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-688577
```

```
Query Match      86.3%; Score 16.4; DB 4; Length 680;
Best Local Similarity 94.4%; Pred. No. 1.6e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 CTTTACTTCATAGCTTTT 18
      ||||| ||||| ||||| |||||
Db      476 CTTTCTTCATAGCTTTT 459
```

```
RESULT 63
US-09-925-065A-688578/c
; Sequence 688578, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR FILING DATE: 2000-10-24
; PRIOR FILING DATE: 2000-11-20
; PRIOR FILING DATE: 2000-11-30
; PRIOR FILING DATE: 2001-01-16
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 688578
; LENGTH: 680
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-688578
```

```
Query Match      86.3%; Score 16.4; DB 4; Length 680;
Best Local Similarity 94.4%; Pred. No. 1.6e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 CTTTACTTCATAGCTTTT 18
      ||||| ||||| ||||| |||||
Db      476 CTTTCTTCATAGCTTTT 459
```

```
RESULT 64
US-09-925-065A-688579/c
; Sequence 688579, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR FILING DATE: 2000-10-24
; PRIOR FILING DATE: 2000-11-20
; PRIOR FILING DATE: 2000-11-30
; PRIOR FILING DATE: 2001-01-16
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 688579
; LENGTH: 680
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-688579
```

```
Query Match      86.3%; Score 16.4; DB 4; Length 680;
Best Local Similarity 94.4%; Pred. No. 1.6e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 CTTTACTTCATAGCTTTT 18
      ||||| ||||| ||||| |||||
Db      476 CTTTCTTCATAGCTTTT 459
```

```
RESULT 65
US-09-925-065A-688580/c
; Sequence 688580, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR FILING DATE: 2000-10-24
; PRIOR FILING DATE: 2000-11-20
; PRIOR FILING DATE: 2000-11-30
; PRIOR FILING DATE: 2001-01-16
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 688580
; LENGTH: 680
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-688580
```

```
Query Match      86.3%; Score 16.4; DB 4; Length 680;
Best Local Similarity 94.4%; Pred. No. 1.6e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 CTTTACTTCATAGCTTTT 18
```

```
Db          476 CTTTCTTCATAGTCTTT 459
||||| ||||| ||||| ||||| |||||
Query Match      86.3%; Score 16.4; DB 5; Length 680;
Best Local Similarity 94.4%; Pred. No. 1.6e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy          1 CTTTACTTCATAGTCTTT 18
||||| ||||| ||||| ||||| |||||
Db          476 CTTTCTTCATAGTCTTT 459
||||| ||||| ||||| ||||| |||||

RESULT 66
US-09-925-065A-688576/c
; Sequence 688576, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925.065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 688576
; LENGTH: 680
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-688576

Query Match      86.3%; Score 16.4; DB 5; Length 680;
Best Local Similarity 94.4%; Pred. No. 1.6e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy          1 CTTTACTTCATAGTCTTT 18
||||| ||||| ||||| ||||| |||||
Db          476 CTTTCTTCATAGTCTTT 459
||||| ||||| ||||| ||||| |||||

RESULT 67
US-09-925-065A-688577/c
; Sequence 688577, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925.065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 688577
; LENGTH: 680
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-688577

Query Match      86.3%; Score 16.4; DB 5; Length 680;
Best Local Similarity 94.4%; Pred. No. 1.6e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy          1 CTTTACTTCATAGTCTTT 18
||||| ||||| ||||| ||||| |||||
Db          476 CTTTCTTCATAGTCTTT 459
||||| ||||| ||||| ||||| |||||

RESULT 68
US-09-925-065A-688578/c
; Sequence 688578, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925.065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 688578
; LENGTH: 680
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-688578

Query Match      86.3%; Score 16.4; DB 5; Length 680;
Best Local Similarity 94.4%; Pred. No. 1.6e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy          1 CTTTACTTCATAGTCTTT 18
||||| ||||| ||||| ||||| |||||
Db          476 CTTTCTTCATAGTCTTT 459
||||| ||||| ||||| ||||| |||||

RESULT 69
US-09-925-065A-688579/c
; Sequence 688579, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925.065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 688579
; LENGTH: 680
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-688579
```

Query Match 86.3%; Score 16.4; DB 5; Length 680;  
Best Local Similarity 94.4%; Pred. No. 1.6e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCTTT 18  
|||||  
Db 476 CTTTCTTCATAGTCTTT 459

## RESULT 70

US-09-925-065A-688580/c  
; Sequence 688580, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 688580  
; LENGTH: 680  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-688580

Query Match 86.3%; Score 16.4; DB 5; Length 680;  
Best Local Similarity 94.4%; Pred. No. 1.6e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCTTT 18  
|||||  
Db 476 CTTTCTTCATAGTCTTT 459

## RESULT 71

US-09-925-065A-75210/c  
; Sequence 75210, Application US/09925065A  
; Publication No. US20040181048A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 75210  
; LENGTH: 708

TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-75210

Query Match 86.3%; Score 16.4; DB 4; Length 708;  
Best Local Similarity 94.4%; Pred. No. 1.6e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TTTTACTTCATAGTCTTTG 19  
|||||  
Db 201 TTTTCTTCATAGTCTTTG 184

## RESULT 72

US-09-925-065A-75211/c  
; Sequence 75211, Application US/09925065A  
; Publication No. US20040181048A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 75211  
; LENGTH: 708  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-75211

Query Match 86.3%; Score 16.4; DB 4; Length 708;  
Best Local Similarity 94.4%; Pred. No. 1.6e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TTTTACTTCATAGTCTTTG 19  
|||||  
Db 201 TTTTCTTCATAGTCTTTG 184

## RESULT 73

US-09-925-065A-75210/c  
; Sequence 75210, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 75210
; LENGTH: 708
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-75210

Query Match      86.3%; Score 16.4; DB 5; Length 708;
Best Local Similarity 94.4%; Pred. No. 1.6e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy  2 TTTACTTCATAGTCTTTG 19
    ||||| ||||| ||||| |||||
Db  201 TTTATTTTCATAGTCTTTG 184

RESULT 74
US-09-925-065A-75211/c
; Sequence 75211, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 75211
; LENGTH: 708
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-75211

Query Match      86.3%; Score 16.4; DB 5; Length 708;
Best Local Similarity 94.4%; Pred. No. 1.6e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy  2 TTTACTTCATAGTCTTTG 19
    ||||| ||||| ||||| |||||
Db  201 TTTATTTTCATAGTCTTTG 184

RESULT 75
US-10-301-480-176449/c
; Sequence 176449, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; PRIOR FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 176449
; LENGTH: 708
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-301-480-176449

Query Match      86.3%; Score 16.4; DB 12; Length 708;
Best Local Similarity 94.4%; Pred. No. 1.6e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy  2 TTTACTTCATAGTCTTTG 19
    ||||| ||||| ||||| |||||
Db  201 TTTATTTTCATAGTCTTTG 184

RESULT 76
US-10-301-480-176450/c
; Sequence 176450, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; PRIOR FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 176450
; LENGTH: 708
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-301-480-176450

Query Match      86.3%; Score 16.4; DB 12; Length 708;
Best Local Similarity 94.4%; Pred. No. 1.6e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy  2 TTTACTTCATAGTCTTTG 19
    ||||| ||||| ||||| |||||
Db  201 TTTATTTTCATAGTCTTTG 184

RESULT 77
US-10-301-480-789858/c
; Sequence 789858, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; PRIOR FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 789858
; LENGTH: 708
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-301-480-789858

Query Match      86.3%; Score 16.4; DB 12; Length 708;
Best Local Similarity 94.4%; Pred. No. 1.6e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy  2 TTTACTTCATAGTCTTTG 19
    ||||| ||||| ||||| |||||
Db  201 TTTATTTTCATAGTCTTTG 184
```

```
Db      201 TTTATTTTCATAGTCTTTG 184
||||| ||||| ||||| ||||| |||||
RESULT 78
US-10-301-480-789859/c
; Sequence 789859, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms
; TITLE OF INVENTION: in the Human Genome
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; PRIOR FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 789859
; LENGTH: 708
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-789859
Query Match      86.3%; Score 16.4; DB 12; Length 708;
Best Local Similarity 94.4%; Pred. No. 1.6e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 TTTACTTCATAGTCTTTG 19
||||| ||||| ||||| ||||| |||||
Db      201 TTTATTTTCATAGTCTTTG 184

RESULT 79
US-09-925-065A-549225/c
; Sequence 549225, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 549225
; LENGTH: 802
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-549225
Query Match      86.3%; Score 16.4; DB 4; Length 802;
Best Local Similarity 94.4%; Pred. No. 1.6e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 TTTACTTCATAGTCTTTG 19
||||| ||||| ||||| ||||| |||||
Db      744 TTTACTTCAAAGTCTTTG 727

RESULT 80
US-09-925-065A-549225/c
; Sequence 549225, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 549225
; LENGTH: 802
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-549225
Query Match      86.3%; Score 16.4; DB 5; Length 802;
Best Local Similarity 94.4%; Pred. No. 1.6e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 TTTACTTCATAGTCTTTG 19
||||| ||||| ||||| ||||| |||||
Db      744 TTTACTTCAAAGTCTTTG 727

RESULT 81
US-10-301-480-527175/c
; Sequence 527175, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 527175
; LENGTH: 802
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-527175
Query Match      86.3%; Score 16.4; DB 12; Length 802;
Best Local Similarity 94.4%; Pred. No. 1.6e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 TTTACTTCATAGTCTTTG 19
||||| ||||| ||||| ||||| |||||
Db      744 TTTACTTCAAAGTCTTTG 727

RESULT 82
US-10-301-480-1140584/c
```

; Sequence 1140584, Application US/10301480  
; Publication No. US20060057564A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms  
; FILE REFERENCE: 108827.137  
; CURRENT APPLICATION NUMBER: US/10/301,480  
; CURRENT FILING DATE: 2002-11-21  
; PRIOR FILING DATE: 2002-08-09  
; PRIOR APPLICATION NUMBER: US 10/215,598  
; PRIOR FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: US 60/311,695  
; NUMBER OF SEQ ID NOS: 1226818  
; SOFTWARE: Fast-SEQ for Windows Version 4.0  
; SEQ ID NO 1140584  
; LENGTH: 802  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-10-301-480-1140584

Query Match 86.3%; Score 16.4; DB 12; Length 802;  
Best Local Similarity 94.4%; Pred. No. 1.6e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TTTACTTCATAGTCTTTG 19  
|||||  
Db 744 TTTACTTCAAGCTTTG 727

RESULT 83  
US-10-424-599-43257  
; Sequence 43257, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with  
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 43257  
; LENGTH: 895  
; TYPE: DNA  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_13905C.1  
US-10-424-599-43257

Query Match 86.3%; Score 16.4; DB 8; Length 895;  
Best Local Similarity 94.4%; Pred. No. 1.6e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCTTT 18  
|||||  
Db 124 CTTTCTTCATAGTCTTT 141

RESULT 84  
US-10-437-963-65932  
; Sequence 65932, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.

; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 65932  
; LENGTH: 995  
; TYPE: DNA  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_66931C.1  
US-10-437-963-65932

Query Match 86.3%; Score 16.4; DB 8; Length 995;  
Best Local Similarity 94.4%; Pred. No. 1.6e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCTTT 18  
|||||  
Db 439 CTTTGCTTCATAGTCTTT 456

RESULT 85  
US-10-301-480-584099  
; Sequence 584099, Application US/10301480  
; Publication No. US20060057564A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms  
; FILE OF INVENTION: in the Human Genome  
; FILE REFERENCE: 108827.137  
; CURRENT APPLICATION NUMBER: US/10/301,480  
; CURRENT FILING DATE: 2002-11-21  
; PRIOR FILING DATE: 2002-08-09  
; PRIOR APPLICATION NUMBER: US 10/215,598  
; PRIOR FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: US 60/311,695  
; NUMBER OF SEQ ID NOS: 1226818  
; SOFTWARE: Fast-SEQ for Windows Version 4.0  
; SEQ ID NO 584099  
; LENGTH: 999  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-10-301-480-584099

Query Match 86.3%; Score 16.4; DB 12; Length 999;  
Best Local Similarity 94.4%; Pred. No. 1.6e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TTTTACTTCATAGTCTTTG 19  
|||||  
Db 597 TTCACCTTCATAGTCTTTG 614

RESULT 86  
US-10-301-480-1197508  
; Sequence 1197508, Application US/10301480  
; Publication No. US20060057564A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms  
; FILE OF INVENTION: in the Human Genome  
; FILE REFERENCE: 108827.137  
; CURRENT APPLICATION NUMBER: US/10/301,480  
; CURRENT FILING DATE: 2002-11-21  
; PRIOR FILING DATE: 2002-08-09  
; PRIOR APPLICATION NUMBER: US 10/215,598  
; PRIOR FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: US 60/311,695  
; NUMBER OF SEQ ID NOS: 1226818

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1197508
; LENGTH: 999
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-1197508

Query Match      86.3%; Score 16.4; DB 12; Length 999;
Best Local Similarity 94.4%; Pred. No. 1.6e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TTTACTTCATAGTCTTTG 19
   |||||||
Db 597 TTCACCTTCATAGTCTTTG 614

RESULT 87
US-09-925-065A-49510/c
; Sequence 49510, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 49510
; LENGTH: 2548
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-49510

Query Match      86.3%; Score 16.4; DB 5; Length 2548;
Best Local Similarity 94.4%; Pred. No. 1.8e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TTTACTTCATAGTCTTTG 19
   |||||||
Db 408 TTTACTTCATATCTTTG 391

RESULT 89
US-10-301-480-150748/c
; Sequence 150748, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; PRIOR FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 150748
; LENGTH: 2548
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-150748

Query Match      86.3%; Score 16.4; DB 12; Length 2548;
Best Local Similarity 94.4%; Pred. No. 1.8e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TTTACTTCATAGTCTTTG 19
   |||||||
Db 408 TTTACTTCATATCTTTG 391

RESULT 90
US-10-301-480-764157/c
; Sequence 764157, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; PRIOR FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 764157
; LENGTH: 2548
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-764157
```



Query Match 86.3%; Score 16.4; DB 12; Length 2548;  
Best Local Similarity 94.4%; Pred. No. 1.8e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TTTACTTCATAGTCTTTG 19  
| | | | | | | | | | | | | | | | | | | | | |  
Db 408 TTTACTTCATATTCCTTTG 391

## RESULT 91

US-10-750-185-53364/c  
; Sequence 53364, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10750,185  
; PRIOR FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 53364  
; LENGTH: 3106  
; TYPE: DNA  
; ORGANISM: Bovine 198668808989808  
US-10-750-185-53364

Query Match 86.3%; Score 16.4; DB 10; Length 3106;  
Best Local Similarity 94.4%; Pred. No. 1.9e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TTTACTTCATAGTCTTTG 19  
| | | | | | | | | | | | | | | | | | | | | |  
Db 305 TCTACTTCATAGTCTTTG 288

## RESULT 92

US-10-750-623-53364/c  
; Sequence 53364, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10750,623  
; PRIOR FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 53364  
; LENGTH: 3106  
; TYPE: DNA  
; ORGANISM: Bovine 198668808989808  
US-10-750-623-53364

Query Match 86.3%; Score 16.4; DB 10; Length 3106;  
Best Local Similarity 94.4%; Pred. No. 1.9e+03;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 2 TTTACTTCATAGTCTTTG 19  
| | | | | | | | | | | | | | | | | | | | | |  
Db 305 TCTACTTCATAGTCTTTG 288

## RESULT 93

US-11-083-784-470117/c  
; Sequence 470117, Application US/11083784  
; Publication No. US20050245475A1  
; GENERAL INFORMATION:  
; APPLICANT: Dharmacon, Inc.  
; APPLICANT: Khvorova, Anastasia  
; APPLICANT: Reynolds, Angela  
; APPLICANT: Leake, Devin  
; APPLICANT: Marshall, William  
; APPLICANT: Scaringe, Stephen  
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
; FILE REFERENCE: 13499US  
; CURRENT APPLICATION NUMBER: US/11/083,784  
; CURRENT FILING DATE: 2005-03-18  
; PRIOR FILING DATE: 2003-11-14  
; PRIOR APPLICATION NUMBER: 60/502,050  
; PRIOR FILING DATE: 2003-09-10  
; PRIOR APPLICATION NUMBER: 60/426,137  
; PRIOR FILING DATE: 2002-11-14  
; NUMBER OF SEQ ID NOS: 1591911  
; SOFTWARE: Proprietary  
; SEQ ID NO 470117  
; LENGTH: 19  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-11-083-784-470117

Query Match 84.2%; Score 16; DB 14; Length 19;  
Best Local Similarity 100.0%; Pred. No. 1.5e+03;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TTTACTTCATAGTCTT 17  
| | | | | | | | | | | | | | | | | | | | | |  
Db 18 TTTACTTCATAGTCTT 3

## RESULT 94

US-11-083-784-470217/c  
; Sequence 470217, Application US/11083784  
; Publication No. US20050245475A1  
; GENERAL INFORMATION:  
; APPLICANT: Dharmacon, Inc.  
; APPLICANT: Khvorova, Anastasia  
; APPLICANT: Reynolds, Angela  
; APPLICANT: Leake, Devin  
; APPLICANT: Marshall, William  
; APPLICANT: Scaringe, Stephen  
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
; FILE REFERENCE: 13499US  
; CURRENT APPLICATION NUMBER: US/11/083,784  
; CURRENT FILING DATE: 2005-03-18  
; PRIOR FILING DATE: 2003-11-14  
; PRIOR APPLICATION NUMBER: 60/502,050  
; PRIOR FILING DATE: 2003-09-10  
; PRIOR APPLICATION NUMBER: 60/426,137  
; PRIOR FILING DATE: 2002-11-14  
; NUMBER OF SEQ ID NOS: 1591911  
; SOFTWARE: Proprietary  
; SEQ ID NO 470217  
; LENGTH: 19  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-11-083-784-470217

Query Match 84.2%; Score 16; DB 14; Length 19;  
Best Local Similarity 100.0%; Pred. No. 1.5e+03;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TTTACTTCATAGTCTT 17  
|||||  
Db 18 TTTACTTCATAGTCTT 3

## RESULT 95

US-11-101-244-470117/c  
; Sequence 470117, Application US/11101244  
; Publication No. US20050246794A1  
; GENERAL INFORMATION:  
; APPLICANT: Dharmakon, Inc.  
; APPLICANT: Khvorova, Anastasia  
; APPLICANT: Reynolds, Angela  
; APPLICANT: Leake, Devin  
; APPLICANT: Marshall, William  
; APPLICANT: Scaringe, Stephen  
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
; FILE REFERENCE: 13499US  
; CURRENT APPLICATION NUMBER: US/11/101,244  
; CURRENT FILING DATE: 2005-04-07  
; PRIOR APPLICATION NUMBER: 60/502,050  
; PRIOR FILING DATE: 2003-09-10  
; PRIOR APPLICATION NUMBER: 60/426,137  
; PRIOR FILING DATE: 2002-11-14  
; NUMBER OF SEQ ID NOS: 1591911  
; SOFTWARE: Proprietary  
; SEQ ID NO 470117  
; LENGTH: 19  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-11-101-244-470117

Query Match 84.2%; Score 16; DB 15; Length 19;  
Best Local Similarity 100.0%; Pred. No. 1.5e+03;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TTTACTTCATAGTCTT 17  
|||||  
Db 18 TTTACTTCATAGTCTT 3

## RESULT 96

US-11-101-244-470217/c  
; Sequence 470217, Application US/11101244  
; Publication No. US20050246794A1  
; GENERAL INFORMATION:  
; APPLICANT: Dharmakon, Inc.  
; APPLICANT: Khvorova, Anastasia  
; APPLICANT: Reynolds, Angela  
; APPLICANT: Leake, Devin  
; APPLICANT: Marshall, William  
; APPLICANT: Scaringe, Stephen  
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA  
; FILE REFERENCE: 13499US  
; CURRENT APPLICATION NUMBER: US/11/101,244  
; CURRENT FILING DATE: 2005-04-07  
; PRIOR APPLICATION NUMBER: 60/502,050  
; PRIOR FILING DATE: 2003-09-10  
; PRIOR APPLICATION NUMBER: 60/426,137  
; PRIOR FILING DATE: 2002-11-14  
; NUMBER OF SEQ ID NOS: 1591911  
; SOFTWARE: Proprietary  
; SEQ ID NO 470217  
; LENGTH: 19  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-11-101-244-470217

Query Match 84.2%; Score 16; DB 15; Length 19;  
Best Local Similarity 100.0%; Pred. No. 1.5e+03;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TTTACTTCATAGTCTT 17  
|||||  
Db 18 TTTACTTCATAGTCTT 3

## RESULT 97

US-11-060-756-43358/c  
; Sequence 43358, Application US/11060756  
; Publication No. US20050221354A1  
; GENERAL INFORMATION:  
; APPLICANT: Wyeth  
; APPLICANT: Mounts, William Martin  
; TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug  
; FILE REFERENCE: AM101083 (031896-042000)  
; CURRENT APPLICATION NUMBER: US/11/060,756  
; CURRENT FILING DATE: 2005-02-18  
; NUMBER OF SEQ ID NOS: 303284  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 43358  
; LENGTH: 25  
; TYPE: DNA  
; ORGANISM: probe  
US-11-060-756-43358

Query Match 84.2%; Score 16; DB 13; Length 25;  
Best Local Similarity 100.0%; Pred. No. 1.6e+03;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCT 16  
|||||  
Db 16 CTTTACTTCATAGTCT 1

## RESULT 98

US-11-060-756-43361/c  
; Sequence 43361, Application US/11060756  
; Publication No. US20050221354A1  
; GENERAL INFORMATION:  
; APPLICANT: Wyeth  
; APPLICANT: Mounts, William Martin  
; TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug  
; FILE REFERENCE: AM101083 (031896-042000)  
; CURRENT APPLICATION NUMBER: US/11/060,756  
; CURRENT FILING DATE: 2005-02-18  
; NUMBER OF SEQ ID NOS: 303284  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 43361  
; LENGTH: 25  
; TYPE: DNA  
; ORGANISM: probe  
US-11-060-756-43361

Query Match 84.2%; Score 16; DB 13; Length 25;  
Best Local Similarity 100.0%; Pred. No. 1.6e+03;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 TACTTCATAGTCTTTG 19  
|||||  
Db 25 TACTTCATAGTCTTTG 10

## RESULT 99

US-11-175-859-39398/c  
; Sequence 39398, Application US/11175859  
; Publication No. US20060024715A1  
; GENERAL INFORMATION:  
; APPLICANT: Affymetrix, Inc.

```
; APPLICANT: Liu, Guoying et al.
; TITLE OF INVENTION: Method of Analysis of Human Polymorphism
; FILE REFERENCE: 3690.1
; CURRENT APPLICATION NUMBER: US/11/175,859
; CURRENT FILING DATE: 2005-07-05
; PRIOR APPLICATION NUMBER: US 60/585,352
; PRIOR FILING DATE: 2004-07-02
; NUMBER OF SEQ ID NOS: 116251
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 39398
; LENGTH: 50
; TYPE: DNA
; ORGANISM: homo sapien
US-11-175-859-39398
```

```
Query Match      84.2%; Score 16; DB 16; Length 50;
Best Local Similarity 88.9%; Pred.No. 1.7e+03;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy      1 CTTTACTTCATAGTCTTT 18
      |||:|||||
Db      30 CATTCCTTCATAGTCTTT 13
```

```
RESULT 100
US-10-301-480-242884
; Sequence 242884, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 242884
; LENGTH: 366
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-242884
```

```
Query Match      84.2%; Score 16; DB 12; Length 366;
Best Local Similarity 88.9%; Pred.No. 2.2e+03;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy      2 TTTACTTCATAGTCTTTG 19
      |||:|||||
Db      44 TTTAYTTCAAAGTCTTTG 61
```

```
Search completed: June 10, 2006, 20:51:02
Job time : 744.7 secs
```

**THIS PAGE IS BLANK**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2006, 15:34:22 ; Search time 9.18333 seconds  
(without alignments)  
262.600 Million cell updates/sec

Title: US-10-600-816-32

Perfect score: 19

Sequence: 1 cttactctagctcttg 19

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 253354 seqs, 63461778 residues

Total number of hits satisfying chosen parameters: 506708

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 1000 summaries

Database : Published Applications\_NA\_New:\*

- 1: /EMC\_Celerra\_SIDS3/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*
- 2: /EMC\_Celerra\_SIDS3/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*
- 3: /EMC\_Celerra\_SIDS3/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*
- 4: /EMC\_Celerra\_SIDS3/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*
- 5: /EMC\_Celerra\_SIDS3/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*
- 6: /EMC\_Celerra\_SIDS3/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*
- 7: /EMC\_Celerra\_SIDS3/ptodata/1/pubpna/US11\_NEW\_PUB.seq:\*
- 8: /EMC\_Celerra\_SIDS3/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	15.4	81.1	522	6	US-10-471-571A-1359
C 2	15.4	81.1	2163	7	US-11-217-529-78062
C 3	15.4	81.1	5376	7	US-11-313-450-17
C 4	15	78.9	1446	6	US-10-953-349-5467
C 5	14.8	77.9	588	7	US-11-183-218-74
C 6	14.8	77.9	654	7	US-11-217-529-76788
C 7	14.8	77.9	676	6	US-10-953-349-6204
C 8	14.8	77.9	956	6	US-10-953-349-2872
C 9	14.8	77.9	1212	7	US-11-217-529-78358
C 10	14.8	77.9	1432	6	US-10-953-349-2801
C 11	14.8	77.9	1882	6	US-10-541-947-5
C 12	14.8	77.9	3551	7	US-11-293-697-610
C 13	14.8	77.9	3816	7	US-11-293-697-662
C 14	14.8	77.9	4218	7	US-11-217-529-75740
C 15	14.4	75.8	285	7	US-11-217-529-11133
C 16	14.4	75.8	285	7	US-11-217-529-80640
C 17	14.4	75.8	432	7	US-11-217-529-475
C 18	14.4	75.8	1457	6	US-10-953-349-11495
C 19	14.4	75.8	3689	7	US-11-293-697-523
C 20	14.2	74.7	25	7	US-11-217-529-161119
C 21	14.2	74.7	25	7	US-11-217-529-172215
C 22	14.2	74.7	192	7	US-11-217-529-166960
C 23	14.2	74.7	255	7	US-11-217-529-80554
C 24	14.2	74.7	517	6	US-10-953-349-10182
C 25	14.2	74.7	606	7	US-11-217-529-82353

771	74.7	14.2	C 26	US-11-217-529-1071	Sequence 1071, Ap
771	74.7	14.2	C 27	US-11-217-529-76887	Sequence 76887, A
819	74.7	14.2	C 28	US-11-217-529-79969	Sequence 79969, A
912	74.7	14.2	C 29	US-10-511-937-501	Sequence 501, App
936	74.7	14.2	C 30	US-10-953-349-5011	Sequence 5011, App
1333	74.7	14.2	C 31	US-10-953-349-7897	Sequence 7897, Ap
1333	74.7	14.2	C 32	US-11-285-701-17	Sequence 17, Appl
1566	74.7	14.2	C 33	US-11-217-529-780	Sequence 780, App
1572	74.7	14.2	C 34	US-11-217-529-77811	Sequence 77811, A
1860	74.7	14.2	C 35	US-11-293-697-1270	Sequence 1270, Ap
1910	74.7	14.2	C 36	US-11-293-697-155	Sequence 155, App
1986	74.7	14.2	C 37	US-10-953-349-11687	Sequence 11687, A
2010	74.7	14.2	C 38	US-11-217-529-2941	Sequence 2941, Ap
2594	74.7	14.2	C 39	US-11-293-697-1439	Sequence 1439, Ap
2628	74.7	14.2	C 40	US-11-217-529-78237	Sequence 78237, A
4675	74.7	14.2	C 41	US-11-327-357-1	Sequence 1, Appli
936	73.7	14	C 42	US-11-217-529-138	Sequence 138, App
1422	73.7	14	C 43	US-10-953-349-2188	Sequence 2188, Ap
2013	73.7	14	C 44	US-10-953-349-8157	Sequence 8157, Ap
25	72.6	13.8	C 45	US-11-217-529-190653	Sequence 190653, A
183	72.6	13.8	C 46	US-10-471-571A-751	Sequence 751, App
272	72.6	13.8	C 47	US-10-473-173-164	Sequence 164, App
330	72.6	13.8	C 48	US-11-301-554-1379	Sequence 1379, Ap
378	72.6	13.8	C 49	US-11-217-529-174589	Sequence 174589, A
564	72.6	13.8	C 50	US-10-488-619-1992	Sequence 1992, Ap
696	72.6	13.8	C 51	US-11-217-529-82746	Sequence 82746, A
729	72.6	13.8	C 52	US-11-217-529-77789	Sequence 77789, A
841	72.6	13.8	C 53	US-10-953-349-8745	Sequence 8745, Ap
867	72.6	13.8	C 54	US-11-217-529-166678	Sequence 166678, A
885	72.6	13.8	C 55	US-10-471-571A-3451	Sequence 3451, Ap
900	72.6	13.8	C 56	US-10-471-571A-3035	Sequence 3035, Ap
946	72.6	13.8	C 57	US-10-953-349-7110	Sequence 7110, Ap
1017	72.6	13.8	C 58	US-10-471-571A-3891	Sequence 3891, Ap
1035	72.6	13.8	C 59	US-11-217-529-78906	Sequence 78906, A
1109	72.6	13.8	C 60	US-10-953-349-14861	Sequence 14861, A
1172	72.6	13.8	C 61	US-10-511-937-441	Sequence 441, App
1191	72.6	13.8	C 62	US-10-471-571A-785	Sequence 785, App
1240	72.6	13.8	C 63	US-10-953-349-20194	Sequence 20194, A
1245	72.6	13.8	C 64	US-11-217-529-78180	Sequence 78180, A
1272	72.6	13.8	C 65	US-10-953-349-1160	Sequence 1160, Ap
1315	72.6	13.8	C 66	US-10-953-349-10559	Sequence 20559, A
1332	72.6	13.8	C 67	US-10-953-349-15758	Sequence 15758, A
1333	72.6	13.8	C 68	US-10-953-349-17823	Sequence 17823, A
1341	72.6	13.8	C 69	US-10-471-571A-4559	Sequence 4559, Ap
1404	72.6	13.8	C 70	US-10-953-349-24131	Sequence 24131, A
1420	72.6	13.8	C 71	US-10-953-349-7778	Sequence 7778, Ap
1461	72.6	13.8	C 72	US-11-217-529-82109	Sequence 82109, A
1485	72.6	13.8	C 73	US-11-217-529-5036	Sequence 5036, Ap
1605	72.6	13.8	C 74	US-11-217-529-82729	Sequence 82729, A
1700	72.6	13.8	C 75	US-11-217-529-421	Sequence 421, App
1804	72.6	13.8	C 76	US-11-217-529-4484	Sequence 4484, Ap
1806	72.6	13.8	C 77	US-10-953-349-20342	Sequence 20342, A
1834	72.6	13.8	C 78	US-10-953-349-1741	Sequence 1741, Ap
1854	72.6	13.8	C 79	US-11-217-529-534	Sequence 534, App
1861	72.6	13.8	C 80	US-10-953-349-18787	Sequence 18787, Ap
1899	72.6	13.8	C 81	US-10-471-571A-4139	Sequence 4139, Ap
2004	72.6	13.8	C 82	US-11-217-529-75539	Sequence 75539, A
2031	72.6	13.8	C 83	US-11-217-529-565	Sequence 565, App
2190	72.6	13.8	C 84	US-10-471-571A-1409	Sequence 1409, Ap
2320	72.6	13.8	C 85	US-10-953-349-8153	Sequence 8153, Ap
2577	72.6	13.8	C 86	US-11-293-697-52	Sequence 52, Appl
2650	72.6	13.8	C 87	US-11-293-697-749	Sequence 749, App
2854	72.6	13.8	C 88	US-11-233-697-2321	Sequence 2321, Ap
3102	72.6	13.8	C 89	US-11-217-529-82302	Sequence 82302, A
3273	72.6	13.8	C 90	US-11-217-529-4879	Sequence 4879, Ap
3447	72.6	13.8	C 91	US-11-217-529-77385	Sequence 77385, A
3528	72.6	13.8	C 92	US-11-217-529-3930	Sequence 3930, Ap
4746	72.6	13.8	C 93	US-10-953-349-2088	Sequence 2088, Ap
5283	72.6	13.8	C 94	US-11-217-529-5476	Sequence 5476, Ap
394191	72.6	13.8	C 95	US-10-506-549-3	Sequence 3, Appl1
70.5	70.5	13.4	C 96	US-11-217-529-74966	Sequence 74966, A
552	70.5	13.4	C 97	US-10-471-571A-1517	Sequence 1517, Ap
878	70.5	13.4	C 98	US-10-953-349-15823	Sequence 15823, A



C 245	12.8	67.4	1059	7	US-11-217-529-82500	Sequence 82500, A	C 318	12.8	67.4	11978	7	US-11-257-851A-63	Sequence 63, Appl
C 246	12.8	67.4	1156	6	US-10-953-349-15088	Sequence 15088, A	C 319	12.6	66.3	25	7	US-11-217-529-88708	Sequence 88708, A
C 247	12.8	67.4	1172	6	US-10-953-349-4182	Sequence 4182, Ap	C 320	12.6	66.3	25	7	US-11-217-529-121670	Sequence 121670, A
C 248	12.8	67.4	1174	6	US-10-953-349-8797	Sequence 8797, Ap	C 321	12.6	66.3	159	6	US-10-471-571A-1507	Sequence 1507, Ap
C 249	12.8	67.4	1200	6	US-10-471-571A-4647	Sequence 4647, Ap	C 322	12.6	66.3	339	7	US-11-217-529-166820	Sequence 166820, A
C 250	12.8	67.4	1224	6	US-10-953-349-9377	Sequence 9377, Ap	C 323	12.6	66.3	342	7	US-11-217-529-78472	Sequence 78472, A
C 251	12.8	67.4	1224	7	US-11-217-529-79666	Sequence 79666, A	C 324	12.6	66.3	384	7	US-11-217-529-75777	Sequence 75777, A
C 252	12.8	67.4	1233	7	US-11-293-697-1566	Sequence 1566, Ap	C 325	12.6	66.3	411	7	US-11-217-529-5764	Sequence 5764, Ap
C 253	12.8	67.4	1249	6	US-10-953-349-14358	Sequence 14358, A	C 326	12.6	66.3	435	7	US-11-217-529-75873	Sequence 75873, A
C 254	12.8	67.4	1254	7	US-11-217-529-77045	Sequence 77045, A	C 327	12.6	66.3	441	7	US-11-217-529-2542	Sequence 2542, Ap
C 255	12.8	67.4	1300	6	US-10-953-349-14532	Sequence 14532, A	C 328	12.6	66.3	468	7	US-11-301-554-604	Sequence 604, App
C 256	12.8	67.4	1306	6	US-10-953-349-31482	Sequence 31482, A	C 329	12.6	66.3	486	7	US-11-217-529-5793	Sequence 5793, Ap
C 257	12.8	67.4	1326	6	US-10-953-349-17594	Sequence 17594, A	C 330	12.6	66.3	507	7	US-11-217-529-190968	Sequence 190968, A
C 258	12.8	67.4	1380	7	US-11-217-529-76526	Sequence 76526, A	C 331	12.6	66.3	510	7	US-11-217-529-3168	Sequence 3168, Ap
C 259	12.8	67.4	1380	7	US-11-217-529-78516	Sequence 78516, A	C 332	12.6	66.3	524	6	US-10-953-349-5669	Sequence 5669, Ap
C 260	12.8	67.4	1431	7	US-11-217-529-1404	Sequence 1404, Ap	C 333	12.6	66.3	536	6	US-10-488-619-1668	Sequence 1668, Ap
C 261	12.8	67.4	1449	7	US-11-217-529-78578	Sequence 78578, A	C 334	12.6	66.3	576	6	US-10-488-619-1882	Sequence 1882, Ap
C 262	12.8	67.4	1470	6	US-10-953-349-82094	Sequence 82094, A	C 335	12.6	66.3	591	7	US-11-217-529-174057	Sequence 174057, A
C 263	12.8	67.4	1470	7	US-11-217-529-3767	Sequence 3767, Ap	C 336	12.6	66.3	603	7	US-11-217-529-765	Sequence 765, App
C 264	12.8	67.4	1473	6	US-10-953-349-24770	Sequence 24770, A	C 337	12.6	66.3	615	7	US-11-217-529-1614	Sequence 1614, Ap
C 265	12.8	67.4	1488	6	US-10-526-429-39	Sequence 39, Appl	C 338	12.6	66.3	624	6	US-10-471-571A-5287	Sequence 5287, Ap
C 266	12.8	67.4	1515	7	US-11-217-529-4555	Sequence 4555, Ap	C 339	12.6	66.3	630	6	US-11-217-529-78824	Sequence 78824, A
C 267	12.8	67.4	1542	7	US-11-217-529-1892	Sequence 1892, Ap	C 340	12.6	66.3	644	6	US-10-196-749-213	Sequence 213, App
C 268	12.8	67.4	1554	7	US-11-217-529-3525	Sequence 3525, Ap	C 341	12.6	66.3	644	7	US-11-101-316-55	Sequence 55, Appl
C 269	12.8	67.4	1581	7	US-11-217-529-82085	Sequence 82085, A	C 342	12.6	66.3	654	7	US-11-217-529-2467	Sequence 2467, Ap
C 270	12.8	67.4	1644	7	US-11-217-529-78032	Sequence 78032, A	C 343	12.6	66.3	658	6	US-10-488-619-2511	Sequence 2511, Ap
C 271	12.8	67.4	1651	6	US-10-953-349-11766	Sequence 11766, A	C 344	12.6	66.3	695	7	US-10-488-619-1881	Sequence 1881, Ap
C 272	12.8	67.4	1653	7	US-11-217-529-3155	Sequence 3155, Ap	C 345	12.6	66.3	717	7	US-11-217-529-81169	Sequence 81169, A
C 273	12.8	67.4	1665	7	US-11-217-529-1913	Sequence 1913, Ap	C 346	12.6	66.3	731	7	US-11-217-529-166792	Sequence 166792, A
C 274	12.8	67.4	1689	7	US-11-217-529-80752	Sequence 80752, A	C 347	12.6	66.3	748	7	US-11-217-529-76004	Sequence 76004, A
C 275	12.8	67.4	1704	6	US-10-471-571A-917	Sequence 917, App	C 348	12.6	66.3	742	6	US-10-511-937-2887	Sequence 2887, Ap
C 276	12.8	67.4	1749	7	US-11-217-529-79614	Sequence 79614, A	C 349	12.6	66.3	752	6	US-10-953-349-16464	Sequence 16464, A
C 277	12.8	67.4	1761	7	US-11-217-529-78138	Sequence 78138, A	C 350	12.6	66.3	771	7	US-11-217-529-1659	Sequence 1659, Ap
C 278	12.8	67.4	1961	7	US-11-293-697-2247	Sequence 2247, Ap	C 351	12.6	66.3	785	6	US-10-953-349-18263	Sequence 18263, A
C 279	12.8	67.4	1983	6	US-10-953-349-5126	Sequence 5126, Ap	C 352	12.6	66.3	792	6	US-10-488-619-2510	Sequence 2510, Ap
C 280	12.8	67.4	2004	7	US-11-217-529-1382	Sequence 1382, Ap	C 353	12.6	66.3	816	6	US-10-953-349-11989	Sequence 11989, A
C 281	12.8	67.4	2031	7	US-11-217-529-78529	Sequence 78529, A	C 354	12.6	66.3	816	7	US-11-217-529-76874	Sequence 76874, A
C 282	12.8	67.4	2088	7	US-11-217-529-586	Sequence 586, App	C 355	12.6	66.3	820	6	US-10-953-349-12479	Sequence 12479, A
C 283	12.8	67.4	2142	7	US-11-217-529-80079	Sequence 80079, A	C 356	12.6	66.3	822	7	US-11-217-529-75872	Sequence 75872, A
C 284	12.8	67.4	2145	7	US-11-217-529-81000	Sequence 81000, A	C 357	12.6	66.3	909	6	US-10-953-349-1444	Sequence 1444, Ap
C 285	12.8	67.4	2158	7	US-11-293-697-3369	Sequence 3369, Ap	C 358	12.6	66.3	915	7	US-11-217-529-1621	Sequence 1621, Ap
C 286	12.8	67.4	2283	7	US-11-217-529-3739	Sequence 3739, Ap	C 359	12.6	66.3	915	7	US-11-217-529-191042	Sequence 191042, A
C 287	12.8	67.4	2298	6	US-10-511-937-525	Sequence 525, App	C 360	12.6	66.3	924	7	US-11-217-529-2811	Sequence 2811, Ap
C 288	12.8	67.4	2514	7	US-11-217-529-5277	Sequence 5277, Ap	C 361	12.6	66.3	924	7	US-11-217-529-173251	Sequence 173251, A
C 289	12.8	67.4	2517	7	US-11-293-697-1130	Sequence 1130, Ap	C 362	12.6	66.3	937	6	US-10-953-349-28474	Sequence 28474, A
C 290	12.8	67.4	2664	7	US-11-217-529-79013	Sequence 79013, A	C 363	12.6	66.3	954	7	US-11-217-529-78390	Sequence 78390, A
C 291	12.8	67.4	2979	7	US-11-253-453-4	Sequence 4, Appl	C 364	12.6	66.3	971	1	US-09-949-525-63	Sequence 63, Appl
C 292	12.8	67.4	2979	7	US-11-178-560-4	Sequence 4, Appl	C 365	12.6	66.3	997	6	US-10-953-349-19221	Sequence 19221, A
C 293	12.8	67.4	3027	6	US-10-471-571A-2401	Sequence 2401, Ap	C 366	12.6	66.3	1003	7	US-11-226-605-16	Sequence 16, Appl
C 294	12.8	67.4	3230	7	US-11-293-697-2392	Sequence 2392, Ap	C 367	12.6	66.3	1008	7	US-11-217-529-81788	Sequence 81788, A
C 295	12.8	67.4	3430	7	US-11-293-697-1442	Sequence 1442, Ap	C 368	12.6	66.3	1008	7	US-11-217-529-191099	Sequence 191099, A
C 296	12.8	67.4	3435	6	US-10-953-349-4556	Sequence 4556, Ap	C 369	12.6	66.3	1011	7	US-11-217-529-77524	Sequence 77524, A
C 297	12.8	67.4	3627	7	US-11-293-697-1352	Sequence 1352, Ap	C 370	12.6	66.3	1026	6	US-10-953-349-5364	Sequence 5364, Ap
C 298	12.8	67.4	3793	7	US-11-293-697-1352	Sequence 1352, Ap	C 371	12.6	66.3	1038	7	US-11-217-529-1880	Sequence 1880, Ap
C 299	12.8	67.4	3843	7	US-11-217-529-1951	Sequence 1951, Ap	C 372	12.6	66.3	1041	7	US-11-217-529-508	Sequence 508, App
C 300	12.8	67.4	3926	7	US-11-251-465-14	Sequence 14, Appl	C 373	12.6	66.3	1041	7	US-11-217-529-82578	Sequence 82578, A
C 301	12.8	67.4	4047	6	US-10-471-571A-3351	Sequence 3351, Ap	C 374	12.6	66.3	1050	6	US-10-953-349-10035	Sequence 10035, A
C 302	12.8	67.4	4321	7	US-11-313-836-7	Sequence 7, Appl	C 375	12.6	66.3	1104	7	US-11-217-529-82147	Sequence 82147, A
C 303	12.8	67.4	4357	6	US-10-511-937-510	Sequence 510, App	C 376	12.6	66.3	1112	6	US-10-953-349-25161	Sequence 25161, A
C 304	12.8	67.4	4434	7	US-11-217-529-76875	Sequence 76875, A	C 377	12.6	66.3	1128	7	US-11-217-529-3541	Sequence 3541, Ap
C 305	12.8	67.4	4473	7	US-11-217-529-1290	Sequence 1290, Ap	C 378	12.6	66.3	1131	7	US-11-217-529-77951	Sequence 77951, A
C 306	12.8	67.4	4576	6	US-10-531-965-1	Sequence 1, Appl	C 379	12.6	66.3	1141	6	US-10-953-349-18899	Sequence 18899, A
C 307	12.8	67.4	4611	7	US-11-217-529-191207	Sequence 191207, A	C 380	12.6	66.3	1141	6	US-10-953-349-22441	Sequence 22441, A
C 308	12.8	67.4	4901	6	US-10-953-349-1429	Sequence 1429, Ap	C 381	12.6	66.3	1149	7	US-11-217-529-76533	Sequence 76533, A
C 309	12.8	67.4	5047	6	US-10-505-928-575	Sequence 575, App	C 382	12.6	66.3	1150	6	US-10-953-349-23523	Sequence 23523, A
C 310	12.8	67.4	5493	7	US-11-217-529-81749	Sequence 81749, A	C 383	12.6	66.3	1161	6	US-10-511-937-3110	Sequence 3110, Ap
C 311	12.8	67.4	5734	6	US-10-505-928-28	Sequence 28, Appl	C 384	12.6	66.3	1161	6	US-10-953-349-4619	Sequence 4619, Ap
C 312	12.8	67.4	7884	7	US-11-217-529-3160	Sequence 3160, Ap	C 385	12.6	66.3	1184	6	US-10-953-349-26589	Sequence 26589, A
C 313	12.8	67.4	9737	1	US-09-484-331-22	Sequence 22, Appl	C 386	12.6	66.3	1188	6	US-10-953-349-27644	Sequence 27644, A
C 314	12.8	67.4	9737	1	US-09-484-331-23	Sequence 23, Appl	C 387	12.6	66.3	1188	7	US-11-217-529-5844	Sequence 5844, Ap
C 315	12.8	67.4	9737	1	US-09-484-331-28	Sequence 28, Appl	C 388	12.6	66.3	1197	6	US-10-953-349-14514	Sequence 14514, A
C 316	12.8	67.4	9871	1	US-09-484-331-24	Sequence 24, Appl	C 389	12.6	66.3	1209	6	US-10-471-571A-3933	Sequence 3933, Ap
C 317	12.8	67.4	10060	1	US-09-484-331-25	Sequence 25, Appl	C 390	12.6	66.3	1214	6	US-10-953-349-19083	Sequence 19083, A





C 537	12.4	65.3	1987	7	US-11-293-697-2224	Sequence 2224, Ap	C 610	12.2	64.2	762	7	US-11-217-529-1258	Sequence 1258, Ap
C 538	12.4	65.3	2067	6	US-10-953-349-5770	Sequence 5770, Ap	C 611	12.2	64.2	774	6	US-10-471-571A-3525	Sequence 3525, Ap
C 539	12.4	65.3	2163	7	US-11-217-529-4479	Sequence 4479, Ap	C 612	12.2	64.2	783	6	US-10-953-349-5500	Sequence 5500, Ap
C 540	12.4	65.3	2249	7	US-11-293-697-722	Sequence 722, App	C 613	12.2	64.2	792	7	US-11-217-529-7801	Sequence 7801, A
C 541	12.4	65.3	2265	7	US-11-217-529-78929	Sequence 78929, Ap	C 614	12.2	64.2	803	6	US-10-488-619-1428	Sequence 1428, Ap
C 542	12.4	65.3	2295	7	US-11-293-697-946	Sequence 946, App	C 615	12.2	64.2	804	6	US-10-953-349-35931	Sequence 35931, A
C 543	12.4	65.3	3126	7	US-11-251-340-4	Sequence 4, Appli	C 616	12.2	64.2	819	6	US-10-471-571A-3739	Sequence 3739, Ap
C 544	12.4	65.3	3296	7	US-11-293-697-688	Sequence 688, App	C 617	12.2	64.2	834	7	US-11-217-529-4582	Sequence 4582, Ap
C 545	12.4	65.3	3495	6	US-10-953-349-39546	Sequence 39546, A	C 618	12.2	64.2	844	6	US-10-953-349-38659	Sequence 38659, A
C 546	12.4	65.3	3667	7	US-11-293-697-33	Sequence 33, Appl	C 619	12.2	64.2	851	6	US-10-953-349-6165	Sequence 6165, Ap
C 547	12.4	65.3	4674	7	US-11-217-529-1182	Sequence 1182, Ap	C 620	12.2	64.2	852	7	US-11-217-529-78919	Sequence 78919, A
C 548	12.2	64.2	25	7	US-11-217-529-24087	Sequence 24087, A	C 621	12.2	64.2	855	7	US-11-217-529-5066	Sequence 5066, Ap
C 549	12.2	64.2	25	7	US-11-217-529-30382	Sequence 30382, A	C 622	12.2	64.2	856	6	US-10-953-349-17527	Sequence 17527, A
C 550	12.2	64.2	25	7	US-11-217-529-94099	Sequence 94099, A	C 623	12.2	64.2	857	6	US-10-953-349-32928	Sequence 32928, A
C 551	12.2	64.2	25	7	US-11-217-529-107138	Sequence 107138, A	C 624	12.2	64.2	858	7	US-11-217-529-173833	Sequence 173833, A
C 552	12.2	64.2	25	7	US-11-217-529-174769	Sequence 174769, A	C 625	12.2	64.2	867	7	US-11-217-529-2211	Sequence 2211, Ap
C 553	12.2	64.2	25	7	US-11-217-529-174771	Sequence 174771, A	C 626	12.2	64.2	885	6	US-10-953-349-10460	Sequence 10460, A
C 554	12.2	64.2	195	6	US-10-554-711-195	Sequence 195, App	C 627	12.2	64.2	885	7	US-11-217-529-77274	Sequence 77274, A
C 555	12.2	64.2	51	7	US-11-245-248-281	Sequence 281, App	C 628	12.2	64.2	888	6	US-10-488-619-2356	Sequence 2356, Ap
C 556	12.2	64.2	51	7	US-11-245-248-335	Sequence 335, App	C 629	12.2	64.2	891	7	US-11-217-529-4427	Sequence 4427, Ap
C 557	12.2	64.2	84	7	US-11-217-529-173140	Sequence 173140, A	C 630	12.2	64.2	897	6	US-10-471-571A-3807	Sequence 3807, Ap
C 558	12.2	64.2	110	7	US-11-194-055-241	Sequence 241, App	C 631	12.2	64.2	897	6	US-11-217-529-5409	Sequence 5409, Ap
C 559	12.2	64.2	126	6	US-10-471-571A-3727	Sequence 3727, Ap	C 632	12.2	64.2	901	6	US-10-953-349-18012	Sequence 18012, A
C 560	12.2	64.2	186	7	US-11-217-529-191000	Sequence 191000, A	C 633	12.2	64.2	906	6	US-10-471-571A-3187	Sequence 3187, Ap
C 561	12.2	64.2	216	6	US-10-471-571A-3415	Sequence 3415, Ap	C 634	12.2	64.2	914	6	US-10-953-349-15932	Sequence 15932, A
C 562	12.2	64.2	237	6	US-10-471-571A-4727	Sequence 4727, Ap	C 635	12.2	64.2	921	7	US-11-217-529-76096	Sequence 76096, A
C 563	12.2	64.2	294	6	US-10-471-571A-2869	Sequence 2869, Ap	C 636	12.2	64.2	925	6	US-10-953-349-39602	Sequence 39602, A
C 564	12.2	64.2	309	7	US-11-301-554-1222	Sequence 1222, Ap	C 637	12.2	64.2	933	6	US-10-953-349-37077	Sequence 37077, A
C 565	12.2	64.2	312	7	US-11-217-529-166750	Sequence 166750, A	C 638	12.2	64.2	942	7	US-11-217-529-82682	Sequence 82682, A
C 566	12.2	64.2	321	7	US-11-301-554-476	Sequence 476, App	C 639	12.2	64.2	951	6	US-10-471-571A-973	Sequence 973, App
C 567	12.2	64.2	324	7	US-11-217-529-173586	Sequence 173586, A	C 640	12.2	64.2	960	6	US-10-953-349-13653	Sequence 13653, A
C 568	12.2	64.2	333	7	US-11-217-529-1041	Sequence 1041, Ap	C 641	12.2	64.2	963	7	US-11-217-529-80262	Sequence 80262, A
C 569	12.2	64.2	345	7	US-11-217-529-191075	Sequence 191075, A	C 642	12.2	64.2	972	6	US-10-471-571A-2991	Sequence 2991, Ap
C 570	12.2	64.2	347	6	US-10-488-619-634	Sequence 634, App	C 643	12.2	64.2	975	7	US-11-217-529-80594	Sequence 80594, A
C 571	12.2	64.2	347	7	US-11-222-810-1	Sequence 1, Appli	C 644	12.2	64.2	978	7	US-11-217-529-258	Sequence 258, App
C 572	12.2	64.2	347	7	US-11-222-810-3	Sequence 3, Appli	C 645	12.2	64.2	978	7	US-11-217-529-75366	Sequence 75366, A
C 573	12.2	64.2	363	7	US-11-217-529-4078	Sequence 4078, Ap	C 646	12.2	64.2	978	7	US-11-217-529-81375	Sequence 81375, A
C 574	12.2	64.2	375	7	US-11-217-529-81887	Sequence 81887, A	C 647	12.2	64.2	984	6	US-10-471-571A-1451	Sequence 1451, Ap
C 575	12.2	64.2	381	6	US-10-471-571A-2657	Sequence 2657, Ap	C 648	12.2	64.2	990	7	US-11-217-529-6230	Sequence 6230, Ap
C 576	12.2	64.2	390	7	US-11-217-529-173480	Sequence 173480, A	C 649	12.2	64.2	990	7	US-11-217-529-77231	Sequence 77231, A
C 577	12.2	64.2	402	7	US-11-217-529-77048	Sequence 77048, A	C 650	12.2	64.2	993	6	US-10-471-571A-2969	Sequence 2969, Ap
C 578	12.2	64.2	435	7	US-11-217-529-174533	Sequence 174533, A	C 651	12.2	64.2	1005	7	US-11-217-529-4665	Sequence 4665, Ap
C 579	12.2	64.2	438	6	US-10-471-571A-3115	Sequence 3115, Ap	C 652	12.2	64.2	1014	7	US-11-217-529-1646	Sequence 1646, Ap
C 580	12.2	64.2	444	6	US-10-953-349-15942	Sequence 15942, A	C 653	12.2	64.2	1038	7	US-11-217-529-77823	Sequence 77823, A
C 581	12.2	64.2	450	7	US-11-217-529-5154	Sequence 5154, Ap	C 654	12.2	64.2	1042	6	US-10-953-349-39084	Sequence 39084, A
C 582	12.2	64.2	460	7	US-11-301-554-709	Sequence 709, App	C 655	12.2	64.2	1042	7	US-11-217-529-79320	Sequence 79320, A
C 583	12.2	64.2	468	7	US-11-301-554-539	Sequence 539, App	C 656	12.2	64.2	1059	6	US-10-953-349-32351	Sequence 32351, A
C 584	12.2	64.2	493	6	US-10-953-349-10668	Sequence 10668, A	C 657	12.2	64.2	1062	7	US-11-217-529-5471	Sequence 5471, Ap
C 585	12.2	64.2	501	7	US-11-183-218-19	Sequence 19, Appl	C 658	12.2	64.2	1062	7	US-11-217-529-81237	Sequence 81237, Ap
C 586	12.2	64.2	506	6	US-10-488-619-2998	Sequence 2998, Ap	C 659	12.2	64.2	1075	6	US-10-953-349-11726	Sequence 11726, A
C 587	12.2	64.2	540	6	US-10-471-571A-637	Sequence 637, App	C 660	12.2	64.2	1098	7	US-11-217-529-5465	Sequence 5465, Ap
C 588	12.2	64.2	558	6	US-10-488-619-2802	Sequence 2802, Ap	C 661	12.2	64.2	1098	7	US-11-217-529-77660	Sequence 77660, A
C 589	12.2	64.2	561	7	US-11-301-554-1341	Sequence 1341, Ap	C 662	12.2	64.2	1110	6	US-10-471-571A-5027	Sequence 5027, Ap
C 590	12.2	64.2	570	6	US-10-471-571A-4957	Sequence 4957, Ap	C 663	12.2	64.2	1110	7	US-11-217-529-4050	Sequence 4050, Ap
C 591	12.2	64.2	578	6	US-10-488-619-1135	Sequence 1135, Ap	C 664	12.2	64.2	1125	6	US-11-217-529-76126	Sequence 76126, A
C 592	12.2	64.2	584	7	US-11-217-529-77047	Sequence 77047, A	C 665	12.2	64.2	1129	6	US-10-953-349-22496	Sequence 22496, A
C 593	12.2	64.2	594	7	US-11-217-529-81783	Sequence 81783, A	C 666	12.2	64.2	1143	7	US-11-217-529-173391	Sequence 173391, A
C 594	12.2	64.2	606	7	US-11-217-529-5255	Sequence 5255, Ap	C 667	12.2	64.2	1146	7	US-11-217-529-190364	Sequence 190364, A
C 595	12.2	64.2	624	6	US-10-471-571A-439	Sequence 439, App	C 668	12.2	64.2	1149	7	US-11-217-529-78468	Sequence 78468, A
C 596	12.2	64.2	632	6	US-10-953-349-37785	Sequence 37785, A	C 669	12.2	64.2	1152	7	US-11-217-529-190877	Sequence 190877, A
C 597	12.2	64.2	632	6	US-10-488-619-2166	Sequence 2166, Ap	C 670	12.2	64.2	1156	6	US-10-953-349-27985	Sequence 27985, A
C 598	12.2	64.2	634	6	US-10-537-455-3	Sequence 3, Appli	C 671	12.2	64.2	1164	7	US-11-217-529-497	Sequence 497, App
C 599	12.2	64.2	638	6	US-10-488-619-2957	Sequence 2957, Ap	C 672	12.2	64.2	1165	6	US-10-953-349-23368	Sequence 23368, A
C 600	12.2	64.2	643	6	US-10-953-349-16589	Sequence 16589, A	C 673	12.2	64.2	1165	6	US-10-953-349-10445	Sequence 10445, A
C 601	12.2	64.2	659	6	US-10-953-349-13084	Sequence 13084, A	C 674	12.2	64.2	1169	6	US-10-953-349-38273	Sequence 38273, A
C 602	12.2	64.2	682	6	US-11-217-529-79096	Sequence 79096, A	C 675	12.2	64.2	1179	6	US-10-953-349-18399	Sequence 18399, A
C 603	12.2	64.2	687	6	US-10-488-619-2853	Sequence 2853, Ap	C 676	12.2	64.2	1182	7	US-11-217-529-2391	Sequence 2391, Ap
C 604	12.2	64.2	693	6	US-11-217-529-174078	Sequence 174078, A	C 677	12.2	64.2	1185	7	US-11-217-529-76810	Sequence 76810, A
C 605	12.2	64.2	708	7	US-11-217-529-76479	Sequence 76479, Ap	C 678	12.2	64.2	1193	6	Sequence 445, App	
C 606	12.2	64.2	714	6	US-10-953-349-1623	Sequence 1623, A	C 679	12.2	64.2	1194	6	Sequence 661, App	
C 607	12.2	64.2	716	6	US-11-217-529-1101	Sequence 1101, Ap	C 680	12.2	64.2	1202	6	Sequence 38715, A	
C 608	12.2	64.2	717	7	US-11-217-529-8502	Sequence 8502, Ap	C 681	12.2	64.2	1203	7	Sequence 78331, A	
C 609	12.2	64.2	735	6			C 682	12.2	64.2	1208	6	Sequence 19298, A	

C 683	12.2	64.2	12.2	6	US-10-953-349-9539	Sequence 9539, Ap	C 756	12.2	64.2	1635	6	US-10-953-349-17109	Sequence 17109, A
C 684	12.2	64.2	12.2	7	US-11-217-529-78198	Sequence 78198, A	C 757	12.2	64.2	1636	6	US-10-953-349-7460	Sequence 7460, Ap
C 685	12.2	64.2	12.2	7	US-11-217-529-174520	Sequence 174520, A	C 758	12.2	64.2	1642	6	US-10-953-349-10772	Sequence 10772, A
C 686	12.2	64.2	12.2	6	US-10-953-349-257	Sequence 257, App	C 759	12.2	64.2	1644	7	US-11-217-529-78032	Sequence 78032, A
C 687	12.2	64.2	12.2	6	US-10-953-349-23687	Sequence 23687, A	C 760	12.2	64.2	1648	6	US-10-953-349-20044	Sequence 20044, A
C 688	12.2	64.2	12.2	6	US-10-953-349-14393	Sequence 14393, A	C 761	12.2	64.2	1649	6	US-10-953-349-8332	Sequence 8332, Ap
C 689	12.2	64.2	12.2	6	US-11-217-529-4464	Sequence 4464, Ap	C 762	12.2	64.2	1653	6	US-10-468-193-25	Sequence 25, Appl
C 690	12.2	64.2	12.2	6	US-11-217-529-76815	Sequence 76815, A	C 763	12.2	64.2	1658	6	US-10-953-349-10166	Sequence 10166, A
C 691	12.2	64.2	12.2	6	US-10-471-571A-2247	Sequence 2247, Ap	C 764	12.2	64.2	1670	7	US-11-174-000-13	Sequence 13, Appl
C 692	12.2	64.2	12.2	6	US-10-953-349-18509	Sequence 18509, A	C 765	12.2	64.2	1681	6	US-10-953-349-13345	Sequence 13345, A
C 693	12.2	64.2	12.2	6	US-11-217-529-76792	Sequence 76792, A	C 766	12.2	64.2	1695	6	US-10-953-349-23231	Sequence 23231, A
C 694	12.2	64.2	12.2	7	US-11-217-529-81314	Sequence 81314, A	C 767	12.2	64.2	1695	7	US-11-217-529-2442	Sequence 2442, Ap
C 695	12.2	64.2	12.2	6	US-10-471-571A-4349	Sequence 4349, Ap	C 768	12.2	64.2	1695	7	US-11-217-529-3778	Sequence 3778, Ap
C 696	12.2	64.2	12.2	6	US-11-217-529-78073	Sequence 78073, A	C 769	12.2	64.2	1701	7	US-11-217-529-190862	Sequence 190862, A
C 697	12.2	64.2	12.2	6	US-10-471-571A-5437	Sequence 5437, Ap	C 770	12.2	64.2	1707	7	US-11-217-529-80360	Sequence 80360, A
C 698	12.2	64.2	12.2	6	US-11-217-529-79276	Sequence 79276, A	C 771	12.2	64.2	1711	6	US-10-953-349-34840	Sequence 34840, A
C 699	12.2	64.2	12.2	7	US-11-217-529-79725	Sequence 79725, A	C 772	12.2	64.2	1731	6	US-10-953-349-23745	Sequence 23745, A
C 700	12.2	64.2	12.2	6	US-10-953-349-3395	Sequence 3395, Ap	C 773	12.2	64.2	1733	6	US-10-953-349-13571	Sequence 13571, A
C 701	12.2	64.2	12.2	6	US-10-471-571A-3241	Sequence 3241, Ap	C 774	12.2	64.2	1734	7	US-11-217-529-76042	Sequence 76042, A
C 702	12.2	64.2	12.2	6	US-10-953-349-14661	Sequence 14661, A	C 775	12.2	64.2	1737	6	US-10-953-349-20636	Sequence 20636, A
C 703	12.2	64.2	12.2	6	US-10-953-349-11380	Sequence 11380, A	C 776	12.2	64.2	1737	7	US-11-217-529-697	Sequence 697, App
C 704	12.2	64.2	12.2	6	US-10-953-349-18646	Sequence 18646, A	C 777	12.2	64.2	1746	7	US-11-217-529-1841	Sequence 1841, Ap
C 705	12.2	64.2	12.2	6	US-11-217-529-78903	Sequence 78903, A	C 778	12.2	64.2	1746	7	US-11-217-529-78093	Sequence 78093, A
C 706	12.2	64.2	12.2	6	US-10-953-349-24513	Sequence 24513, A	C 779	12.2	64.2	1824	6	US-10-953-349-1041	Sequence 1041, Ap
C 707	12.2	64.2	12.2	7	US-11-217-529-77527	Sequence 77527, A	C 780	12.2	64.2	1824	7	US-11-217-529-77881	Sequence 77881, A
C 708	12.2	64.2	12.2	6	US-10-953-349-832	Sequence 832, App	C 781	12.2	64.2	1832	6	US-10-953-349-12617	Sequence 12617, A
C 709	12.2	64.2	12.2	6	US-10-953-349-21995	Sequence 21995, A	C 782	12.2	64.2	1848	6	US-10-524-648-25	Sequence 25, Appl
C 710	12.2	64.2	12.2	6	US-10-471-571A-5313	Sequence 5313, Ap	C 783	12.2	64.2	1851	7	US-11-217-529-3685	Sequence 3685, Ap
C 711	12.2	64.2	12.2	7	US-11-217-529-80001	Sequence 80001, A	C 784	12.2	64.2	1851	7	US-11-217-529-80356	Sequence 80356, Ap
C 712	12.2	64.2	12.2	6	US-10-953-349-14176	Sequence 14176, A	C 785	12.2	64.2	1859	7	US-11-217-529-80537	Sequence 80537, A
C 713	12.2	64.2	12.2	6	US-10-953-349-20566	Sequence 20566, A	C 786	12.2	64.2	1890	7	US-11-217-529-77994	Sequence 77994, A
C 714	12.2	64.2	12.2	6	US-10-953-349-23800	Sequence 23800, A	C 787	12.2	64.2	1898	6	US-10-524-648-27	Sequence 27, Appl
C 715	12.2	64.2	12.2	6	US-10-471-571A-1945	Sequence 1945, Ap	C 788	12.2	64.2	1901	7	US-11-293-697-1169	Sequence 1169, Ap
C 716	12.2	64.2	12.2	6	US-11-217-529-3441	Sequence 3441, Ap	C 789	12.2	64.2	1915	6	US-10-953-349-13595	Sequence 13595, A
C 717	12.2	64.2	12.2	6	US-10-953-349-22542	Sequence 22542, A	C 790	12.2	64.2	1948	7	US-11-293-697-265	Sequence 265, App
C 718	12.2	64.2	12.2	6	US-11-217-529-80260	Sequence 80260, A	C 791	12.2	64.2	1970	6	US-10-953-349-10697	Sequence 10697, A
C 719	12.2	64.2	12.2	7	US-11-217-529-78077	Sequence 78077, A	C 792	12.2	64.2	1975	6	US-11-313-836-53	Sequence 53, Appl
C 720	12.2	64.2	12.2	6	US-10-471-571A-3678	Sequence 3678, Ap	C 793	12.2	64.2	1979	6	US-10-953-349-860	Sequence 860, App
C 721	12.2	64.2	12.2	6	US-11-217-529-191161	Sequence 191161, A	C 794	12.2	64.2	1983	7	US-11-217-529-1680	Sequence 1680, Ap
C 722	12.2	64.2	12.2	7	US-11-217-529-79577	Sequence 79577, A	C 795	12.2	64.2	2004	7	US-11-217-529-1799	Sequence 179, App
C 723	12.2	64.2	12.2	7	US-11-217-529-78925	Sequence 78925, A	C 796	12.2	64.2	2037	6	US-10-196-749-591	Sequence 591, App
C 724	12.2	64.2	12.2	7	US-11-222-810-11	Sequence 11, Appl	C 797	12.2	64.2	2042	6	US-10-953-349-4433	Sequence 4433, Ap
C 725	12.2	64.2	12.2	7	US-11-222-810-13	Sequence 13, Appl	C 798	12.2	64.2	2052	6	US-10-471-571A-3983	Sequence 3983, Ap
C 726	12.2	64.2	12.2	6	US-10-953-349-3678	Sequence 3678, Ap	C 799	12.2	64.2	2052	7	US-11-293-697-2258	Sequence 2258, Ap
C 727	12.2	64.2	12.2	6	US-11-217-529-576	Sequence 576, App	C 800	12.2	64.2	2064	7	US-11-293-697-1695	Sequence 1695, Ap
C 728	12.2	64.2	12.2	6	US-10-953-349-28502	Sequence 28502, A	C 801	12.2	64.2	2072	6	US-10-953-349-11127	Sequence 11127, A
C 729	12.2	64.2	12.2	6	US-10-953-349-20753	Sequence 20753, A	C 802	12.2	64.2	2082	7	US-11-217-529-2079	Sequence 2079, Ap
C 730	12.2	64.2	12.2	6	US-10-953-349-4603	Sequence 4603, Ap	C 803	12.2	64.2	2091	7	US-11-217-529-77094	Sequence 77094, A
C 731	12.2	64.2	12.2	6	US-11-217-529-81729	Sequence 81729, A	C 804	12.2	64.2	2103	6	US-10-953-349-24505	Sequence 24505, A
C 732	12.2	64.2	12.2	6	US-10-953-349-28518	Sequence 28518, A	C 805	12.2	64.2	2121	7	US-11-136-524-59	Sequence 59, Appl
C 733	12.2	64.2	12.2	6	US-10-953-349-36486	Sequence 36486, A	C 806	12.2	64.2	2126	7	US-11-293-697-1258	Sequence 1258, Ap
C 734	12.2	64.2	12.2	6	US-11-217-529-1368	Sequence 1368, Ap	C 807	12.2	64.2	2126	7	US-11-293-697-2457	Sequence 2457, Ap
C 735	12.2	64.2	12.2	7	US-11-217-529-76116	Sequence 76116, A	C 808	12.2	64.2	2133	6	US-10-471-571A-7367	Sequence 7367, Ap
C 736	12.2	64.2	12.2	7	US-11-217-529-77310	Sequence 77310, A	C 809	12.2	64.2	2154	7	US-11-217-529-77607	Sequence 77607, A
C 737	12.2	64.2	12.2	7	US-11-217-529-75890	Sequence 75890, A	C 810	12.2	64.2	2154	7	US-11-217-529-77607	Sequence 77607, A
C 738	12.2	64.2	12.2	7	US-11-217-529-76036	Sequence 76036, A	C 811	12.2	64.2	2167	7	US-11-293-697-111	Sequence 111, App
C 739	12.2	64.2	12.2	6	US-10-953-349-2135	Sequence 2135, Ap	C 812	12.2	64.2	2171	7	US-11-293-697-50	Sequence 50, Appl
C 740	12.2	64.2	12.2	6	US-10-953-349-21156	Sequence 21156, A	C 813	12.2	64.2	2187	7	US-11-217-529-4024	Sequence 4024, Ap
C 741	12.2	64.2	12.2	6	US-11-217-529-78235	Sequence 78235, A	C 814	12.2	64.2	2214	6	US-10-953-349-34043	Sequence 34043, A
C 742	12.2	64.2	12.2	6	US-10-953-349-15946	Sequence 15946, A	C 815	12.2	64.2	2234	1	US-09-949-925-52	Sequence 52, Appl
C 743	12.2	64.2	12.2	6	US-10-953-349-2192	Sequence 2192, Ap	C 816	12.2	64.2	2247	7	US-11-217-529-1781	Sequence 1781, Ap
C 744	12.2	64.2	12.2	6	US-10-471-571A-1815	Sequence 1815, Ap	C 817	12.2	64.2	2247	7	US-11-217-529-76537	Sequence 76537, A
C 745	12.2	64.2	12.2	6	US-10-953-349-609	Sequence 609, App	C 818	12.2	64.2	2248	6	US-11-293-697-66	Sequence 66, Appl
C 746	12.2	64.2	12.2	6	US-11-217-529-77941	Sequence 77941, A	C 819	12.2	64.2	2248	6	US-10-953-349-8336	Sequence 8336, Ap
C 747	12.2	64.2	12.2	6	US-11-217-529-132	Sequence 132, App	C 820	12.2	64.2	2250	6	US-10-953-349-6743	Sequence 6743, Ap
C 748	12.2	64.2	12.2	6	US-10-953-349-9484	Sequence 9484, Ap	C 821	12.2	64.2	2261	7	US-11-293-697-1861	Sequence 1861, Ap
C 749	12.2	64.2	12.2	6	US-11-217-529-1197	Sequence 1197, Ap	C 822	12.2	64.2	2262	7	US-11-217-529-306	Sequence 306, App
C 750	12.2	64.2	12.2	6	US-10-953-349-1904	Sequence 1904, Ap	C 823	12.2	64.2	2268	7	US-11-217-529-344	Sequence 344, App
C 751	12.2	64.2	12.2	6	US-10-953-349-9886	Sequence 9886, Ap	C 824	12.2	64.2	2271	6	US-10-953-349-11004	Sequence 11004, A
C 752	12.2	64.2	12.2	6	US-10-953-349-10672	Sequence 10672, A	C 825	12.2	64.2	2306	6	US-10-953-349-37942	Sequence 37942, A
C 753	12.2	64.2	12.2	6	US-10-953-349-24024	Sequence 24024, A	C 826	12.2	64.2	2313	6	US-11-217-529-77342	Sequence 77342, A
C 754	12.2	64.2	12.2	7	US-11-217-529-326	Sequence 326, App	C 827	12.2	64.2	2323	7	US-11-293-697-2027	Sequence 2027, Ap
C 755	12.2	64.2	12.2	7	US-11-217-529-706	Sequence 706, App	C 828	12.2	64.2	2334	7	US-11-217-529-810	Sequence 810, App

C 829	12.2	64.2	2375	6	US-10-953-349-11035	Sequence 11035, A	C 902	12.2	64.2	5957	6	US-10-473-173-106	Sequence 106, App
C 830	12.2	64.2	2387	7	US-11-217-529-11035	Sequence 50, Appl	C 903	12.2	64.2	6045	7	US-11-217-529-1312	Sequence 1312, Ap
C 831	12.2	64.2	2436	7	US-11-217-529-1652	Sequence 1522, Ap	C 904	12.2	64.2	6317	7	US-10-511-937-387	Sequence 387, App
C 832	12.2	64.2	2436	7	US-11-217-529-82283	Sequence 82283, A	C 905	12.2	64.2	6891	7	US-11-251-724-2	Sequence 2, Appl
C 833	12.2	64.2	2440	7	US-11-293-697-1135	Sequence 1135, Ap	C 906	12.2	64.2	7131	7	US-11-217-529-77074	Sequence 77074, A
C 834	12.2	64.2	2479	7	US-11-293-697-1966	Sequence 1966, Ap	C 907	12.2	64.2	7254	7	US-11-217-529-1173	Sequence 1173, A
C 835	12.2	64.2	2498	7	US-11-293-697-1092	Sequence 1092, Ap	C 908	12.2	64.2	7873	6	US-10-485-397-7	Sequence 7, Appl
C 836	12.2	64.2	2537	6	US-10-505-928-149	Sequence 149, App	C 909	12.2	64.2	7943	6	US-10-485-397-8	Sequence 8, Appl
C 837	12.2	64.2	2559	7	US-11-293-697-83	Sequence 83, Appl	C 910	12.2	64.2	8090	7	US-11-226-605-86	Sequence 86, Appl
C 838	12.2	64.2	2562	7	US-11-217-529-4523	Sequence 4523, A	C 911	12.2	64.2	8488	7	US-11-145-307A-166	Sequence 166, App
C 839	12.2	64.2	2568	7	US-11-217-529-77262	Sequence 77262, Ap	C 912	12.2	64.2	8488	7	US-10-480-962-38	Sequence 38, Appl
C 840	12.2	64.2	2572	1	US-09-949-925-36	Sequence 36, Appl	C 913	12.2	64.2	9449	6	US-10-480-962-39	Sequence 39, Appl
C 841	12.2	64.2	2607	7	US-11-217-529-3100	Sequence 3100, Ap	C 914	12.2	64.2	12342	7	US-11-217-529-4644	Sequence 4644, Ap
C 842	12.2	64.2	2616	7	US-11-217-529-3604	Sequence 3604, Ap	C 915	12.2	64.2	15304	6	US-10-505-928-828	Sequence 828, App
C 843	12.2	64.2	2638	7	US-11-293-697-879	Sequence 879, App	C 916	12.2	64.2	37436	6	US-10-473-173-32	Sequence 32, Appl
C 844	12.2	64.2	2645	7	US-11-181-115-25	Sequence 25, Appl	C 917	12.2	64.2	56580	6	US-10-553-298-1	Sequence 1, Appl
C 845	12.2	64.2	2648	7	US-11-293-697-807	Sequence 807, App	C 918	12.2	64.2	70665	6	US-10-505-928-596	Sequence 596, App
C 846	12.2	64.2	2650	7	US-11-293-697-241	Sequence 241, App	C 919	12.2	64.2	70665	6	US-10-505-928-596	Sequence 596, App
C 847	12.2	64.2	2655	6	US-10-473-173-96	Sequence 96, Appl	C 920	12.2	64.2	118899	7	US-11-189-279-64	Sequence 64, Appl
C 848	12.2	64.2	2663	6	US-10-953-349-8793	Sequence 8793, Ap	C 921	12.2	64.2	135090	6	US-10-505-928-607	Sequence 607, App
C 849	12.2	64.2	2682	7	US-11-217-529-571	Sequence 571, App	C 922	12.2	64.2	135090	6	US-10-505-928-607	Sequence 607, App
C 850	12.2	64.2	2694	7	US-11-217-529-5531	Sequence 5531, Ap	C 923	12.2	64.2	394191	6	US-10-506-549-3	Sequence 3, Appl
C 851	12.2	64.2	2733	7	US-11-217-529-76516	Sequence 76516, A	C 924	12.2	64.2	394191	6	US-11-217-529-137480	Sequence 137480, A
C 852	12.2	64.2	2760	7	US-11-217-529-80783	Sequence 80783, A	C 925	12.2	63.2	25	7	US-11-217-529-142134	Sequence 142134, A
C 853	12.2	64.2	2778	7	US-11-217-529-1916	Sequence 1916, Ap	C 926	12.2	63.2	631	7	US-11-301-554-566	Sequence 566, App
C 854	12.2	64.2	2808	7	US-11-181-115-23	Sequence 23, Appl	C 927	12.2	63.2	751	6	US-10-953-349-31486	Sequence 31486, A
C 855	12.2	64.2	2808	7	US-11-181-115-27	Sequence 27, Appl	C 928	12.2	63.2	1172	6	US-10-953-349-6628	Sequence 6628, Ap
C 856	12.2	64.2	2808	7	US-11-181-115-41	Sequence 41, Appl	C 929	12.2	63.2	1177	6	US-10-953-349-8634	Sequence 8634, Ap
C 857	12.2	64.2	2811	7	US-11-217-529-3053	Sequence 3053, Ap	C 930	12.2	63.2	1204	6	US-10-953-349-29986	Sequence 29986, A
C 858	12.2	64.2	2886	7	US-11-217-529-75510	Sequence 75510, A	C 931	12.2	63.2	1254	7	US-11-217-529-81803	Sequence 81803, A
C 859	12.2	64.2	2901	7	US-11-217-529-2168	Sequence 2168, Ap	C 932	12.2	63.2	1284	7	US-11-217-529-1514	Sequence 1514, A
C 860	12.2	64.2	2916	7	US-11-217-529-81609	Sequence 81609, Ap	C 933	12.2	63.2	1284	7	US-11-217-529-77448	Sequence 77448, A
C 861	12.2	64.2	2976	6	US-10-953-349-37847	Sequence 37847, A	C 934	12.2	63.2	1454	6	US-10-953-349-2764	Sequence 2764, Ap
C 862	12.2	64.2	2980	7	US-11-293-697-465	Sequence 465, App	C 935	12.2	63.2	1515	6	US-11-217-529-3477	Sequence 3477, Ap
C 863	12.2	64.2	2983	7	US-11-293-697-589	Sequence 589, App	C 936	12.2	63.2	1620	6	US-10-471-571A-4207	Sequence 4207, Ap
C 864	12.2	64.2	2988	7	US-11-217-529-80377	Sequence 80377, A	C 937	12.2	63.2	1631	6	US-10-953-349-12559	Sequence 12559, A
C 865	12.2	64.2	2997	7	US-11-217-529-1582	Sequence 1582, Ap	C 938	12.2	63.2	1725	7	US-11-217-529-81352	Sequence 81352, A
C 866	12.2	64.2	3012	7	US-11-217-529-3912	Sequence 3912, Ap	C 939	12.2	63.2	1737	7	US-11-217-529-1440	Sequence 1440, Ap
C 867	12.2	64.2	3012	7	US-11-293-697-720	Sequence 720, App	C 940	12.2	63.2	1937	1	US-09-949-925-61	Sequence 61, Appl
C 868	12.2	64.2	3045	7	US-11-293-697-1338	Sequence 1338, Ap	C 941	12.2	63.2	2049	7	US-11-217-529-79820	Sequence 79820, A
C 869	12.2	64.2	3069	7	US-11-217-529-80037	Sequence 80027, A	C 942	12.2	63.2	2073	7	US-11-217-529-352	Sequence 352, App
C 870	12.2	64.2	3128	7	US-11-293-697-1259	Sequence 1259, Ap	C 943	12.2	63.2	2076	7	US-11-106-014-9	Sequence 9, Appl
C 871	12.2	64.2	3198	6	US-10-953-349-8411	Sequence 8411, Ap	C 944	12.2	63.2	2082	7	US-11-217-529-2079	Sequence 2079, Ap
C 872	12.2	64.2	3252	7	US-11-318-813-3	Sequence 3, Appl	C 945	12.2	63.2	2139	6	US-10-511-937-338	Sequence 338, App
C 873	12.2	64.2	3271	7	US-11-222-810-22	Sequence 22, Appl	C 946	12.2	63.2	2168	7	US-11-293-697-2011	Sequence 2011, Ap
C 874	12.2	64.2	3271	7	US-11-293-697-244	Sequence 244, App	C 947	12.2	63.2	2847	7	US-11-293-697-403	Sequence 403, App
C 875	12.2	64.2	3278	7	US-11-145-307A-198	Sequence 198, App	C 948	12.2	63.2	3609	7	US-11-217-529-77868	Sequence 77868, A
C 876	12.2	64.2	3399	7	US-11-217-529-73	Sequence 73, Appl	C 949	12.2	63.2	4308	7	US-11-217-529-82409	Sequence 82409, A
C 877	12.2	64.2	3408	7	US-11-217-529-2270	Sequence 2270, Ap	C 950	12.2	63.2	4656	7	US-11-217-529-3130	Sequence 3130, Ap
C 878	12.2	64.2	3435	7	US-11-217-529-989	Sequence 989, App	C 951	12.2	63.2	5748	7	US-11-217-529-3601	Sequence 3601, Ap
C 879	12.2	64.2	3441	7	US-11-217-529-76893	Sequence 76893, A	C 952	12.2	63.2	5833	7	US-11-024-544A-139	Sequence 139, App
C 880	12.2	64.2	3451	7	US-11-145-307A-192	Sequence 192, App	C 953	12.2	63.2	5833	7	US-11-024-545-61	Sequence 61, Appl
C 881	12.2	64.2	3486	6	US-10-953-349-16524	Sequence 16524, A	C 954	12.2	63.2	5833	7	US-11-251-466-35	Sequence 35, Appl
C 882	12.2	64.2	3628	15	US-11-145-307A-15	Sequence 15, Appl	C 955	12.2	63.2	5833	7	US-11-254-173-46	Sequence 46, Appl
C 883	12.2	64.2	3855	7	US-11-217-529-79166	Sequence 79166, A	C 956	12.2	63.2	5833	7	US-11-264-784-131	Sequence 131, App
C 884	12.2	64.2	4015	7	US-11-293-697-627	Sequence 627, App	C 957	12.2	63.2	5833	7	US-11-264-737-159	Sequence 159, App
C 885	12.2	64.2	4021	7	US-11-293-697-472	Sequence 472, App	C 958	12.2	63.2	5833	7	US-11-265-761-133	Sequence 133, App
C 886	12.2	64.2	4098	6	US-10-510-508-2	Sequence 2, Appl	C 959	12.2	63.2	6457	7	US-11-264-784-117	Sequence 117, App
C 887	12.2	64.2	4110	6	US-10-471-571A-739	Sequence 739, App	C 960	12.2	63.2	6457	7	US-11-264-737-166	Sequence 166, App
C 888	12.2	64.2	4179	7	US-11-217-529-176	Sequence 176, App	C 961	12.2	63.2	6457	7	US-11-265-761-142	Sequence 142, App
C 889	12.2	64.2	4257	7	US-11-217-529-2743	Sequence 2743, Ap	C 962	12.2	63.2	6498	7	US-11-264-784-148	Sequence 148, App
C 890	12.2	64.2	4419	7	US-11-217-529-2484	Sequence 2484, Ap	C 963	12.2	63.2	6498	7	US-11-264-737-200	Sequence 200, App
C 891	12.2	64.2	4498	7	US-11-217-529-190990	Sequence 190990, A	C 964	12.2	63.2	6498	7	US-11-265-761-163	Sequence 163, App
C 892	12.2	64.2	4654	6	US-10-505-928-20	Sequence 20, Appl	C 965	12.2	63.2	6540	7	US-11-254-173-60	Sequence 60, Appl
C 893	12.2	64.2	4688	7	US-11-293-697-222	Sequence 222, App	C 966	12.2	63.2	6540	7	US-11-264-784-135	Sequence 135, App
C 894	12.2	64.2	4761	7	US-11-217-529-1821	Sequence 1821, Ap	C 967	12.2	63.2	6540	7	US-11-264-737-187	Sequence 187, App
C 895	12.2	64.2	4804	7	US-11-181-115-42	Sequence 42, Appl	C 968	12.2	63.2	6540	7	US-11-265-761-139	Sequence 139, App
C 896	12.2	64.2	4963	7	US-11-217-529-76536	Sequence 76536, A	C 969	12.2	63.2	6561	7	US-11-264-784-147	Sequence 147, App
C 897	12.2	64.2	5009	7	US-11-244-452-8	Sequence 8, Appl	C 970	12.2	63.2	6561	7	US-11-264-784-150	Sequence 150, App
C 898	12.2	64.2	5072	6	US-10-953-349-7998	Sequence 7998, Ap	C 971	12.2	63.2	6561	7	US-11-264-737-199	Sequence 199, App
C 899	12.2	64.2	5540	1	US-09-970-076-5	Sequence 5, Appl	C 972	12.2	63.2	6561	7	US-11-264-737-202	Sequence 202, App
C 900	12.2	64.2	5688	7	US-11-217-529-76708	Sequence 76708, A	C 973	12.2	63.2	6561	7	US-11-265-761-162	Sequence 162, App
C 901	12.2	64.2	5920	7	US-11-145-307A-185	Sequence 185, App	C 974	12.2	63.2	6561	7	US-11-265-761-165	Sequence 165, App

Sequence 57, Appl  
Sequence 133, App  
Sequence 185, App  
Sequence 138, App  
Sequence 168, App  
Sequence 88, Appl  
Sequence 53, Appl  
Sequence 61, Appl  
Sequence 140, Appl  
Sequence 192, App  
Sequence 155, App  
Sequence 144, App  
Sequence 153, App  
Sequence 205, App  
Sequence 168, App  
Sequence 119, App  
Sequence 167, App  
Sequence 144, App  
Sequence 154, App  
Sequence 206, App  
Sequence 169, App  
Sequence 41, Appl  
Sequence 138, App  
Sequence 190, App  
Sequence 153, App  
Sequence 67, Appl

12 63.2 6912 7 US-11-254-173-57  
12 63.2 6912 7 US-11-264-784-133  
12 63.2 6912 7 US-11-264-737-185  
12 63.2 6912 7 US-11-265-761-138  
12 63.2 7323 7 US-11-024-544A-168  
12 63.2 7323 7 US-11-024-544A-68  
12 63.2 7323 7 US-11-251-466-53  
12 63.2 7323 7 US-11-254-173-61  
12 63.2 7323 7 US-11-264-784-140  
12 63.2 7323 7 US-11-264-737-192  
12 63.2 7323 7 US-11-265-761-155  
12 63.2 7822 7 US-11-024-544A-144  
12 63.2 7822 7 US-11-264-784-153  
12 63.2 7822 7 US-11-264-737-205  
12 63.2 7822 7 US-11-265-761-168  
12 63.2 7879 7 US-11-264-784-119  
12 63.2 7879 7 US-11-264-737-167  
12 63.2 7879 7 US-11-265-761-144  
12 63.2 7930 7 US-11-264-784-154  
12 63.2 7930 7 US-11-264-737-206  
12 63.2 7930 7 US-11-265-761-169  
12 63.2 8015 7 US-11-251-466-41  
12 63.2 8015 7 US-11-264-784-138  
12 63.2 8015 7 US-11-264-737-190  
12 63.2 8015 7 US-11-265-761-153  
1000 12 63.2 8084 7 US-11-024-545-67

ALIGNMENTS

RESULT 1  
US-10-471-571A-1359/c  
; Sequence 1359, Application US/10471571A  
; Publication No. US20060115490A1  
; GENERAL INFORMATION:  
; APPLICANT: CHIRON Spa  
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS PROTEINS AND NUCLEIC ACIDS  
; FILE REFERENCE: P026927WO  
; CURRENT APPLICATION NUMBER: US/10/471,571A  
; CURRENT FILING DATE: 2003-09-12  
; PRIOR APPLICATION NUMBER: GB-0107661.1  
; PRIOR FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 5642  
; SOFTWARE: SeqWin99, version 1.03  
; SEQ ID NO 1359  
; LENGTH: 522  
; TYPE: DNA  
; ORGANISM: Staphylococcus aureus  
US-10-471-571A-1359

Query Match 81.1%; Score 15.4; DB 6; Length 522;  
Best Local Similarity 94.1%; Pred. No. 42;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TTACTTCATAGTCTTT 18  
| | | | | | | | | | | | | | | | | |  
Db 209 TTACTTCATAGCCTTT 193

RESULT 2  
US-11-217-529-78062/c  
; Sequence 78062, Application US/11217529  
; Publication No. US20060099612A1  
; GENERAL INFORMATION:  
; APPLICANT: SUNTORY LIMITED  
; APPLICANT: NAKAO, YOSHIHIRO  
; APPLICANT: NAKAMURA, NORIHISA  
; APPLICANT: KODAMA, YUKIKO  
; APPLICANT: FUJIMURA, TOMOKO  
; APPLICANT: ASHIKARI, TOSHIKO  
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS  
; FILE REFERENCE: S-38-285

; CURRENT APPLICATION NUMBER: US/11/217,529  
; CURRENT FILING DATE: 2005-09-02  
; PRIOR APPLICATION NUMBER: US 10/932,182  
; PRIOR FILING DATE: 2004-09-02  
; NUMBER OF SEQ ID NOS: 197023  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 78062  
; LENGTH: 2163  
; TYPE: DNA  
; ORGANISM: Saccharomyces pastorianus  
US-11-217-529-78062

Query Match 81.1%; Score 15.4; DB 7; Length 2163;  
Best Local Similarity 94.1%; Pred. No. 49;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 TTACTTCATAGTCTTTG 19  
| | | | | | | | | | | | | | | | | |  
Db 1662 TTCTTCATAGTCTTTG 1646

RESULT 3  
US-11-313-450-17  
; Sequence 17, Application US/11313450  
; Publication No. US20060110778A1  
; GENERAL INFORMATION:  
; APPLICANT: Adorante, Joeseeph S.  
; APPLICANT: Ehring, George R.  
; TITLE OF INVENTION: High-Throughput Screens For Identifying  
; TITLE OF INVENTION: Selective Persistent Sodium Channels Channel Blockers  
; FILE REFERENCE: 17407 CIP (AP)  
; CURRENT APPLICATION NUMBER: US/11/313,450  
; CURRENT FILING DATE: 2005-12-19  
; PRIOR APPLICATION NUMBER: 09/989797  
; PRIOR FILING DATE: 2001-11-20  
; PRIOR APPLICATION NUMBER: 60/252771  
; PRIOR FILING DATE: 2000-11-22  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 17  
; LENGTH: 5376  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-313-450-17

Query Match 81.1%; Score 15.4; DB 7; Length 5376;  
Best Local Similarity 94.1%; Pred. No. 55;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TTACTTCATAGTCTTT 18  
| | | | | | | | | | | | | | | | | |  
Db 3857 TTACTTCGTAGTCTTT 3873

RESULT 4  
US-10-953-349-5467/c  
; Sequence 5467, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 5467  
; LENGTH: 1446  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-10-953-349-5467

Query Match 78.9%; Score 15; DB 6; Length 1446;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 TACTTCATAGCTTTT 18  
|||||  
Db 1187 TACTTCATAGCTTTT 1173

## RESULT 5

US-11-183-218-74/c  
; Sequence 74, Application US/11183218  
; Publication No. US2006008906A1  
; GENERAL INFORMATION:  
; APPLICANT: Neose Technologies, Inc.  
; APPLICANT: DePree, Shawn  
; APPLICANT: Zopf, David  
; APPLICANT: Bayer, Robert  
; APPLICANT: Hakes, David  
; APPLICANT: Chen, Xi  
; APPLICANT: Bowe, Caryne  
; TITLE OF INVENTION: ERYTHROPOIETIN: REMODELING AND  
; TITLE OF INVENTION: GLYCOCONJUGATION OF ERYTHROPOIETIN  
; FILE REFERENCE: 040853-01-5083-US02  
; CURRENT APPLICATION NUMBER: US/11/183.218  
; CURRENT FILING DATE: 2005-07-15  
; PRIOR FILING DATE: 2005-04-09  
; PRIOR APPLICATION NUMBER: PCT/US02/32263  
; PRIOR FILING DATE: 2002-10-09  
; PRIOR APPLICATION NUMBER: US 60/407,527  
; PRIOR FILING DATE: 2002-08-28  
; PRIOR APPLICATION NUMBER: US 60/404,249  
; PRIOR FILING DATE: 2002-08-16  
; PRIOR APPLICATION NUMBER: US 60/396,594  
; PRIOR FILING DATE: 2002-07-17  
; PRIOR APPLICATION NUMBER: US 60/391,777  
; PRIOR FILING DATE: 2002-06-25  
; PRIOR APPLICATION NUMBER: US 60/387,292  
; PRIOR FILING DATE: 2002-06-07  
; PRIOR APPLICATION NUMBER: US 60/344,692  
; PRIOR FILING DATE: 2001-11-19  
; PRIOR APPLICATION NUMBER: US 60/334,301  
; PRIOR FILING DATE: 2001-11-28  
; PRIOR APPLICATION NUMBER: US 60/334,233  
; PRIOR FILING DATE: 2001-11-28

; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 75  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 74  
; LENGTH: 588  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-183-218-74

Query Match 77.9%; Score 14.8; DB 7; Length 588;  
Best Local Similarity 88.9%; Pred. No. 81;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGCTTTT 18  
|||||  
Db 565 CTTTACTTCATAGCTTTT 548

## RESULT 6

US-11-217-529-76788/c  
; Sequence 76788, Application US/11217529  
; Publication No. US20060099612A1  
; GENERAL INFORMATION:  
; APPLICANT: SUNTORY LIMITED  
; APPLICANT: NAKAO, YOSHIHIRO  
; APPLICANT: NAKAMURA, NORIHISA  
; APPLICANT: KODAMA, YUKIKO

; APPLICANT: FUJIMURA, TOMOKO  
; APPLICANT: ASHIKARI, TOSHIHIKO  
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS  
; FILE REFERENCE: S-38-285  
; CURRENT APPLICATION NUMBER: US/11/217,529  
; CURRENT FILING DATE: 2005-09-02  
; PRIOR APPLICATION NUMBER: US 10/932,182  
; PRIOR FILING DATE: 2004-09-02  
; NUMBER OF SEQ ID NOS: 197023  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 76788  
; LENGTH: 654  
; TYPE: DNA  
; ORGANISM: Saccharomyces pastorianus  
US-11-217-529-76788

Query Match 77.9%; Score 14.8; DB 7; Length 654;  
Best Local Similarity 88.9%; Pred. No. 82;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TTTTACTTCATAGCTTTG 19  
|||||  
Db 578 TTGACTTCATAGCTCTG 561

## RESULT 7

US-10-953-349-6204/c  
; Sequence 6204, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 6204  
; LENGTH: 676  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-10-953-349-6204

Query Match 77.9%; Score 14.8; DB 6; Length 676;  
Best Local Similarity 88.9%; Pred. No. 82;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TTTTACTTCATAGCTTTG 19  
|||||  
Db 199 TTTACTTCATAGCTCTG 182

## RESULT 8

US-10-953-349-2872  
; Sequence 2872, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 2872  
; LENGTH: 956  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-10-953-349-2872

Query Match 77.9%; Score 14.8; DB 6; Length 956;

Best Local Similarity 88.9%; Pred. No. 86;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TTTACTTCATAGTCCTTTG 19  
|||||  
Db 440 TTTACTTCAAAGCTTTTG 457

## RESULT 9

US-11-217-529-78358/c  
; Sequence 78358, Application US/11217529  
; Publication No. US20060099612A1  
; GENERAL INFORMATION:  
; APPLICANT: SUNTORY LIMITED  
; APPLICANT: NAKAO, YOSHIHIRO  
; APPLICANT: NAKAMURA, NORIHISA  
; APPLICANT: KODAMA, YUKIKO  
; APPLICANT: FUJIMURA, TOMOKO  
; APPLICANT: ASHIKARI, TOSHIHIKO  
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS  
; FILE REFERENCE: S-38-285  
; CURRENT APPLICATION NUMBER: US/11/217,529  
; CURRENT FILING DATE: 2005-09-02  
; PRIOR APPLICATION NUMBER: US 10/932,182  
; PRIOR FILING DATE: 2004-09-02  
; NUMBER OF SEQ ID NOS: 197023  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 78358  
; LENGTH: 1212  
; TYPE: DNA  
; ORGANISM: Saccharomyces pastorianus  
US-11-217-529-78358

Query Match 77.9%; Score 14.8; DB 7; Length 1212;  
Best Local Similarity 88.9%; Pred. No. 88;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCCTTT 18  
|||||  
Db 1127 CTTTGCTTCATAGTGTTT 1110

## RESULT 10

US-10-953-349-2801  
; Sequence 2801, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 2801  
; LENGTH: 1432  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-10-953-349-2801

Query Match 77.9%; Score 14.8; DB 6; Length 1432;  
Best Local Similarity 88.9%; Pred. No. 90;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCCTTT 18  
|||||  
Db 57 CTTTCTTCATCGCTTTT 74

## RESULT 11

US-10-541-947-5/c  
; Sequence 5, Application US/10541947

Publication No. US20060095980A1  
; GENERAL INFORMATION:  
; APPLICANT: North Carolina State University  
; APPLICANT: Pettite, James  
; APPLICANT: Pardue, Samuel  
; TITLE OF INVENTION: DEPLETION OF ENDOGENOUS PRIMORDIAL GERM CELLS IN AVIAN SPECIES  
; FILE REFERENCE: 297/204 PCT  
; CURRENT APPLICATION NUMBER: US/10/541,947  
; CURRENT FILING DATE: 2005-07-08  
; PRIOR APPLICATION NUMBER: US 60/440,424  
; PRIOR FILING DATE: 2003-01-16  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 5  
; LENGTH: 1882  
; TYPE: DNA  
; ORGANISM: Gallus gallus  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (174)..(1043)  
US-10-541-947-5

Query Match 77.9%; Score 14.8; DB 6; Length 1882;  
Best Local Similarity 88.3%; Pred. No. 92;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TTTTACTTCATAGTCCTTTG 19  
|||||  
Db 1065 TGTACTTCAAAGCTTTTG 1048

## RESULT 12

US-11-293-697-610  
; Sequence 610, Application US/11293697  
; Publication No. US20060105376A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: Novel full length cDNA  
; FILE REFERENCE: H1-A0106  
; CURRENT APPLICATION NUMBER: US/11/293,697  
; CURRENT FILING DATE: 2005-12-05  
; PRIOR APPLICATION NUMBER: US/10/108,260  
; PRIOR FILING DATE: 2002-03-28  
; NUMBER OF SEQ ID NOS: 5458  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 610  
; LENGTH: 3551  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-293-697-610

Query Match 77.9%; Score 14.8; DB 7; Length 3551;  
Best Local Similarity 88.9%; Pred. No. 99;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCCTTT 18  
|||||  
Db 2987 CTTTAATTCATAGCCTTT 3004

## RESULT 13

US-11-293-697-662  
; Sequence 662, Application US/11293697  
; Publication No. US20060105376A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: Novel full length cDNA  
; FILE REFERENCE: H1-A0106  
; CURRENT APPLICATION NUMBER: US/11/293,697  
; CURRENT FILING DATE: 2005-12-05  
; PRIOR APPLICATION NUMBER: US/10/108,260  
; PRIOR FILING DATE: 2002-03-28  
; NUMBER OF SEQ ID NOS: 5458



```
QY      2 TTTACTTCATAGCTTT 17
      ||| ||| ||| ||| |||
Db      314 TTTCTTCATAGCTTT 329

RESULT 18
US-10-953-349-11495/c
; Sequence 11495, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 11495
; LENGTH: 1457
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-10-953-349-11495

Query Match      75.8%; Score 14.4; DB 6; Length 1457;
Best Local Similarity 93.8%; Pred. No. 1.4e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 TTTACTTCATAGCTTT 17
      ||| ||| ||| ||| |||
Db      753 TTGACTTCATAGCTTT 738

RESULT 19
US-11-293-697-523/c
; Sequence 523, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: HI-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 523
; LENGTH: 3689
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-293-697-523

Query Match      75.8%; Score 14.4; DB 7; Length 3689;
Best Local Similarity 93.8%; Pred. No. 1.5e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      4 TACTTCATAGCTTTTG 19
      ||| ||| ||| ||| |||
Db      3428 TACTTCAAAGTCTTTTG 3413

RESULT 20
US-11-217-529-161319
; Sequence 161319, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
```

```
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 161319
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-161319

Query Match      74.7%; Score 14.2; DB 7; Length 25;
Best Local Similarity 84.2%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 CTTTACTTCATAGCTTTTG 19
      ||| ||| ||| ||| |||
Db      3 CTTTGGTCATCGTCTTTG 21

RESULT 21
US-11-217-529-172215
; Sequence 172215, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 172215
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-172215

Query Match      74.7%; Score 14.2; DB 7; Length 25;
Best Local Similarity 84.2%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 CTTTACTTCATAGCTTTTG 19
      ||| ||| ||| ||| |||
Db      2 CTTTCTTCGTAGTTTGG 20

RESULT 22
US-11-217-529-166960
; Sequence 166960, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
```





Db 690 CTTGACTTCATCGTATTG 672

## RESULT 27

US-11-217-529-76887/c  
; Sequence 76887, Application US/11217529  
; Publication No. US20060099612A1  
; GENERAL INFORMATION:  
; APPLICANT: SUNTORY LIMITED  
; APPLICANT: NAKAO, YOSHIHIRO  
; APPLICANT: NAKAMURA, NORIHISA  
; APPLICANT: KODAMA, YUKIKO  
; APPLICANT: FUJIMURA, TOMOKO  
; APPLICANT: ASHIKARI, TOSHIKO  
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS  
; FILE REFERENCE: S-38-285  
; CURRENT APPLICATION NUMBER: US/11/217,529  
; CURRENT FILING DATE: 2005-09-02  
; PRIOR APPLICATION NUMBER: US 10/932,182  
; PRIOR FILING DATE: 2004-09-02  
; NUMBER OF SEQ ID NOS: 197023  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 76887  
; LENGTH: 771  
; TYPE: DNA  
; ORGANISM: Saccharomyces pastorianus  
US-11-217-529-76887

Query Match 74.7%; Score 14.2; DB 7; Length 771;  
Best Local Similarity 84.2%; Pred. No. 1.6e+02;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGCTTTG 19

Db 690 CTTGACTTCATCGTATTG 672

## RESULT 28

US-11-217-529-79969/c  
; Sequence 79969, Application US/11217529  
; Publication No. US20060099612A1  
; GENERAL INFORMATION:  
; APPLICANT: SUNTORY LIMITED  
; APPLICANT: NAKAO, YOSHIHIRO  
; APPLICANT: NAKAMURA, NORIHISA  
; APPLICANT: KODAMA, YUKIKO  
; APPLICANT: FUJIMURA, TOMOKO  
; APPLICANT: ASHIKARI, TOSHIKO  
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS  
; FILE REFERENCE: S-38-285  
; CURRENT APPLICATION NUMBER: US/11/217,529  
; CURRENT FILING DATE: 2005-09-02  
; PRIOR APPLICATION NUMBER: US 10/932,182  
; PRIOR FILING DATE: 2004-09-02  
; NUMBER OF SEQ ID NOS: 197023  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 79969  
; LENGTH: 819  
; TYPE: DNA  
; ORGANISM: Saccharomyces pastorianus  
US-11-217-529-79969

Query Match 74.7%; Score 14.2; DB 7; Length 819;  
Best Local Similarity 84.2%; Pred. No. 1.6e+02;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGCTTTG 19

Db 738 CTTTCTCTCTCGTCTTTG 720

## RESULT 29

US-10-511-937-501/c

; Sequence 501, Application US/10511937  
; Publication No. US20060088836A1  
; GENERAL INFORMATION:  
; APPLICANT: EXPRESSION DIAGNOSTICS, INC.  
; APPLICANT: Wohlgenuth, Jay  
; APPLICANT: Fry, Kirk  
; APPLICANT: Woodward, Robert  
; APPLICANT: Ly, Ngoc  
; APPLICANT: Prentice, James  
; APPLICANT: Morris, MacDonald  
; APPLICANT: Rosenberg, Steven  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING  
; TITLE OF INVENTION: AND MONITORING TRANSPLANT REJECTION  
; FILE REFERENCE: 506612000104  
; CURRENT APPLICATION NUMBER: US/10/511,937  
; CURRENT FILING DATE: 2004-10-19  
; PRIOR APPLICATION NUMBER: PCT/US2003/012946  
; PRIOR FILING DATE: 2003-04-24  
; PRIOR APPLICATION NUMBER: US 10/131,831  
; PRIOR FILING DATE: 2002-04-24  
; PRIOR APPLICATION NUMBER: US 10/325,899  
; PRIOR FILING DATE: 2002-12-20  
; NUMBER OF SEQ ID NOS: 3117  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 501  
; LENGTH: 912  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-511-937-501

Query Match 74.7%; Score 14.2; DB 6; Length 912;  
Best Local Similarity 84.2%; Pred. No. 1.6e+02;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGCTTTG 19

Db 276 CTTAATTCATAGTTTGT 258

## RESULT 30

US-10-953-349-5011  
; Sequence 5011, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; TITLE OF INVENTION: ENCODED THERBY  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 5011  
; LENGTH: 936  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-10-953-349-5011

Query Match 74.7%; Score 14.2; DB 6; Length 936;  
Best Local Similarity 84.2%; Pred. No. 1.6e+02;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGCTTTG 19

Db 849 CTTTACTTCAGTTTCTTTG 867

## RESULT 31

US-10-953-349-7897  
; Sequence 7897, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nikolai et al.

```
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: ENCODED THERBY
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 7897
; LENGTH: 1333
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-10-953-349-7897

Query Match          74.7%; Score 14.2; DB 6; Length 1333;
Best Local Similarity 84.2%; Pred. No. 1.7e+02;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19
    ||||| ||||| ||||| |||||
Db 96 CTTTCTCTCAAAGGCTTTG 114

RESULT 32
US-11-285-701-17
; Sequence 17, Application US/11285701
; Publication No. US20060117415A1
; GENERAL INFORMATION:
; APPLICANT: Keetman, Ulrich
; APPLICANT: Linemann, Ute
; APPLICANT: Herbers, Karin
; APPLICANT: Hillebrand, Helke
; TITLE OF INVENTION: EXPRESSION CASSETTES FOR ROOT-PREFERENTIAL EXPRESSION IN PLANTS
; FILE REFERENCE: 13173-00018-US
; CURRENT APPLICATION NUMBER: US/11/285,701
; CURRENT FILING DATE: 2005-11-22
; PRIOR APPLICATION NUMBER: EP 04028202.2
; PRIOR FILING DATE: 2004-11-27
; PRIOR APPLICATION NUMBER: EP 05002261.5
; PRIOR FILING DATE: 2005-02-03
; PRIOR APPLICATION NUMBER: EP 05002849.7
; PRIOR FILING DATE: 2005-02-11
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 17
; LENGTH: 1333
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (42)..(1094)
; OTHER INFORMATION: coding for peroxidase ATP9a protein
US-11-285-701-17

Query Match          74.7%; Score 14.2; DB 7; Length 1333;
Best Local Similarity 84.2%; Pred. No. 1.7e+02;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19
    ||||| ||||| ||||| |||||
Db 96 CTTTCTCTCAAAGGCTTTG 114

RESULT 33
US-11-217-529-780/c
; Sequence 780, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
```

```
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 780
; LENGTH: 1566
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-780

Query Match          74.7%; Score 14.2; DB 7; Length 1566;
Best Local Similarity 84.2%; Pred. No. 1.7e+02;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19
    ||||| ||||| ||||| |||||
Db 502 CTTCACTTGATAGTCTTTG 484

RESULT 34
US-11-217-529-77811/c
; Sequence 77811, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 77811
; LENGTH: 1572
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-77811

Query Match          74.7%; Score 14.2; DB 7; Length 1572;
Best Local Similarity 84.2%; Pred. No. 1.7e+02;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19
    ||||| ||||| ||||| |||||
Db 428 CTTTCTCTCATAGTCTTAG 410

RESULT 35
US-11-293-697-1270
; Sequence 1270, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: HI-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1270
; LENGTH: 1860
```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-293-697-1270

Query Match      74.7%; Score 14.2; DB 7; Length 1860;
Best Local Similarity 84.2%; Pred. No. 1.8e+02;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCCTTTG 19
    ||||| ||||| |||||
Db 1410 CTTTGCTTCATAGTAATGG 1428

RESULT 36
US-11-293-697-155
; Sequence 155, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cdna
; FILE REFERENCE: HI-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 155
; LENGTH: 1910
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-293-697-155

Query Match      74.7%; Score 14.2; DB 7; Length 1910;
Best Local Similarity 84.2%; Pred. No. 1.8e+02;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCCTTTG 19
    ||||| ||||| |||||
Db 1652 CTTTACTTAATAGTTTGG 1670

RESULT 37
US-10-953-349-11687
; Sequence 11687, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCES-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 11687
; LENGTH: 1986
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-10-953-349-11687

Query Match      74.7%; Score 14.2; DB 6; Length 1986;
Best Local Similarity 84.2%; Pred. No. 1.8e+02;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCCTTTG 19
    ||||| ||||| |||||
Db 133 CTTCAATTCATATCTTTG 151

RESULT 38
US-11-217-529-2941/c
; Sequence 2941, Application US/11217529
```

```
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHIISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 2941
; LENGTH: 2010
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-2941

Query Match      74.7%; Score 14.2; DB 7; Length 2010;
Best Local Similarity 84.2%; Pred. No. 1.8e+02;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCCTTTG 19
    ||||| ||||| |||||
Db 891 CTTTTCATCAAGTCCTTTG 873

RESULT 39
US-11-293-697-1439/c
; Sequence 1439, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cdna
; FILE REFERENCE: HI-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1439
; LENGTH: 2594
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-293-697-1439

Query Match      74.7%; Score 14.2; DB 7; Length 2594;
Best Local Similarity 84.2%; Pred. No. 1.8e+02;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCCTTTG 19
    ||||| ||||| |||||
Db 1719 CTTTCCTTTATAATCTTTG 1701

RESULT 40
US-11-217-529-78237/c
; Sequence 78237, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHIISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
```

; CURRENT APPLICATION NUMBER: US/11/217,529  
; CURRENT FILING DATE: 2005-09-02  
; PRIOR APPLICATION NUMBER: US 10/932,182  
; PRIOR FILING DATE: 2004-09-02  
; NUMBER OF SEQ ID NOS: 197023  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 78237  
; LENGTH: 2628  
; TYPE: DNA  
; ORGANISM: Saccharomyces pastorianus  
US-11-217-529-78237

Query Match 74.7%; Score 14.2; DB 7; Length 2628;  
Best Local Similarity 84.2%; Pred. No. 1.8e+02;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19  
||| ||||| |||||  
Db 2148 CTTGGCTTCAATAATCTTG 2130

RESULT 41  
US-11-327-357-1/c  
; Sequence 1, Application US/11327357  
; Publication No. US20060105983A1  
; GENERAL INFORMATION:  
; APPLICANT: BTG International Ltd  
; APPLICANT: BEARD DR, PETER  
; APPLICANT: RAJ DR, KENNETH  
; TITLE OF INVENTION: CYTOTOXIC AGENTS  
; FILE REFERENCE: 142184W0  
; CURRENT APPLICATION NUMBER: US/11/327,357  
; CURRENT FILING DATE: 2006-01-09  
; PRIOR APPLICATION NUMBER: 0009887.1  
; PRIOR FILING DATE: 2000-04-20  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 4675  
; TYPE: DNA  
; ORGANISM: adeno-associated virus 2  
US-11-327-357-1

Query Match 74.7%; Score 14.2; DB 7; Length 4675;  
Best Local Similarity 84.2%; Pred. No. 1.9e+02;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTTG 19  
||| ||||| |||||  
Db 3793 CTTTCTTCATGCTCCTTG 3775

RESULT 42  
US-11-217-529-138/c  
; Sequence 138, Application US/11217529  
; Publication No. US20060099612A1  
; GENERAL INFORMATION:  
; APPLICANT: SUNTORY LIMITED  
; APPLICANT: NAKAO, YOSHIHIRO  
; APPLICANT: NAKAMURA, NORIHISA  
; APPLICANT: KODAMA, YUKIKO  
; APPLICANT: FUJIMURA, TOMOKO  
; APPLICANT: ASHKARI, TOSHIKO  
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS  
; FILE REFERENCE: S-38-285  
; CURRENT APPLICATION NUMBER: US/11/217,529  
; CURRENT FILING DATE: 2005-09-02  
; PRIOR APPLICATION NUMBER: US 10/932,182  
; PRIOR FILING DATE: 2004-09-02  
; NUMBER OF SEQ ID NOS: 197023  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 138  
; LENGTH: 936

; TYPE: DNA  
; ORGANISM: Saccharomyces pastorianus  
US-11-217-529-138

Query Match 73.7%; Score 14; DB 7; Length 936;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 CTTCATAGTCTTTG 19  
||| ||||| |||||  
Db 171 CTTCATAGTCTTTG 158

RESULT 43  
US-10-953-349-2188/c  
; Sequence 2188, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nickolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; TITLE OF INVENTION: ENCODED THERBY  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 2188  
; LENGTH: 1422  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-10-953-349-2188

Query Match 73.7%; Score 14; DB 6; Length 1422;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 CTTCATAGTCTTTG 19  
||| ||||| |||||  
Db 803 CTTCATAGTCTTTG 790

RESULT 44  
US-10-953-349-8157/c  
; Sequence 8157, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nickolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; TITLE OF INVENTION: ENCODED THERBY  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 8157  
; LENGTH: 2013  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-10-953-349-8157

Query Match 73.7%; Score 14; DB 6; Length 2013;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 CTTCATAGTCTTTG 19  
||| ||||| |||||  
Db 809 CTTCATAGTCTTTG 796

RESULT 45  
US-11-217-529-190653  
; Sequence 190653, Application US/11217529  
; Publication No. US20060099612A1

```
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHIISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/217,529
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 190653
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-190653

Query Match      72.6%; Score 13.8; DB 7; Length 25;
Best Local Similarity 88.2%; Pred. No. 1.7e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 TTACTTCATAGTCTTTG 19
   ||||| ||||| |||||
Db 4 TTCCTTTATAGTCTTTG 20

RESULT 46
US-10-471-571A-751/c
; Sequence 751, Application US/10471571A
; Publication No. US20060115490A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026927W0
; CURRENT APPLICATION NUMBER: US/10/471,571A
; CURRENT FILING DATE: 2003-09-12
; PRIOR APPLICATION NUMBER: GB-0107661.1
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 5642
; SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 751
; LENGTH: 183
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-10-471-571A-751

Query Match      72.6%; Score 13.8; DB 6; Length 183;
Best Local Similarity 88.2%; Pred. No. 2.1e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TTACTTCATAGTCTTT 18
   ||||| ||||| |||||
Db 83 TTTACTTCATAGTCTTT 67

RESULT 47
US-10-473-173-164/c
; Sequence 164, Application US/10473173
; Publication No. US2006008823A1
; GENERAL INFORMATION:
; APPLICANT: VAN ANDEL INSTITUTE
; TITLE OF INVENTION: Microarray Gene Expression Profiling in Clear Cell Renal Cell
; FILE REFERENCE: 38345-170094
; CURRENT APPLICATION NUMBER: US/10/473,173
; CURRENT FILING DATE: 2003-09-29
; PRIOR APPLICATION NUMBER: US 60/279,411
; PRIOR FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 498
```

```
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 164
; LENGTH: 272
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-473-173-164

Query Match      72.6%; Score 13.8; DB 6; Length 272;
Best Local Similarity 88.2%; Pred. No. 2.2e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 TTACTTCATAGTCTTTG 19
   ||||| ||||| ||||| |||||
Db 79 TTACATTATAGTCTTTG 63

RESULT 48
US-11-301-554-1379
; Sequence 1379, Application US/11301554
; Publication No. US20060088527A1
; GENERAL INFORMATION:
; APPLICANT: Henderson, Robert A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Kalos, Michael D.
; APPLICANT: Sleath, Paul R.
; APPLICANT: Johnson, Jeffrey C.
; APPLICANT: Retter, Marc W.
; APPLICANT: Durham, Margarita
; APPLICANT: Carter, Darrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: McNabb, Andria
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.478C21
; CURRENT APPLICATION NUMBER: US/11/301,554
; CURRENT FILING DATE: 2005-12-13
; PRIOR APPLICATION NUMBER: US 10/283,017
; PRIOR FILING DATE: 2002-10-28
; PRIOR APPLICATION NUMBER: US 10/113,872
; PRIOR FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: US 10/017,754
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 09/902,941
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 09/849,626
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: US 09/736,457
; PRIOR FILING DATE: 2000-12-13
; PRIOR APPLICATION NUMBER: US 09/702,705
; PRIOR FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: US 09/677,419
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: US 09/671,325
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US 09/658,824
; PRIOR FILING DATE: 2000-09-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 2157
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1379
; LENGTH: 330
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-301-554-1379

Query Match      72.6%; Score 13.8; DB 7; Length 330;
Best Local Similarity 88.2%; Pred. No. 2.2e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 TTACTTCATAGTCTTTG 19
```

Db 96 TTACATTATAGTCTTTG 112  
||||| |||||||

## RESULT 49

US-11-217-529-174589  
; Sequence 174589, Application US/11217529  
; Publication No. US20060099612A1  
; GENERAL INFORMATION:  
; APPLICANT: SUNTORY LIMITED  
; APPLICANT: NAKAO, YOSHIHIRO  
; APPLICANT: NAKAMURA, NORIHISA  
; APPLICANT: KODAMA, YUKIKO  
; APPLICANT: FUJIMURA, TOMOKO  
; APPLICANT: ASHIKARI, TOSHIHIKO  
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS  
; FILE REFERENCE: S-38-285  
; CURRENT APPLICATION NUMBER: US/11/217,529  
; CURRENT FILING DATE: 2005-09-02  
; PRIOR FILING DATE: 2004-09-02  
; NUMBER OF SEQ ID NOS: 197023  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 174589  
; LENGTH: 378  
; TYPE: DNA  
; ORGANISM: Saccharomyces pastorianus  
US-11-217-529-174589

Query Match 72.6%; Score 13.8; DB 7; Length 378;  
Best Local Similarity 88.2%; Pred. No. 2.3e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 TTACTTCATAGTCTTTG 19  
||| |||||||  
Db 213 TTCTTTATAGTCTTTG 229

## RESULT 50

US-10-488-619-1992/c  
; Sequence 1992, Application US/10488619  
; Publication No. US20060099578A1  
; GENERAL INFORMATION:  
; APPLICANT: Greenlee, winner and Sullivan, P.C.  
; TITLE OF INVENTION: Human Mitochondrial DNA Polymorphisms, Haplogroups, Associations  
; TITLE OF INVENTION: Physiological Conditions, And Genotyping Arrays  
; FILE REFERENCE: 98-01 WO  
; CURRENT APPLICATION NUMBER: US/10/488,619  
; CURRENT FILING DATE: 2004-03-01  
; NUMBER OF SEQ ID NOS: 3040  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1992  
; LENGTH: 564  
; TYPE: DNA  
; ORGANISM: Mus musculus  
US-10-488-619-1992

Query Match 72.6%; Score 13.8; DB 6; Length 564;  
Best Local Similarity 88.2%; Pred. No. 2.4e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 TTACTTCATAGTCTTTG 19  
||| |||||||  
Db 479 TTCTTCATAGTCTTCG 463

## RESULT 51

US-11-217-529-82746/c  
; Sequence 82746, Application US/11217529  
; Publication No. US20060099612A1  
; GENERAL INFORMATION:  
; \* APPLICANT: SUNTORY LIMITED  
; APPLICANT: NAKAO, YOSHIHIRO

; APPLICANT: NAKAMURA, NORIHISA  
; APPLICANT: KODAMA, YUKIKO  
; APPLICANT: FUJIMURA, TOMOKO  
; APPLICANT: ASHIKARI, TOSHIHIKO  
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS  
; FILE REFERENCE: S-38-285  
; CURRENT APPLICATION NUMBER: US/11/217,529  
; CURRENT FILING DATE: 2005-09-02  
; PRIOR FILING DATE: 2004-09-02  
; NUMBER OF SEQ ID NOS: 197023  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 82746  
; LENGTH: 696  
; TYPE: DNA  
; ORGANISM: Saccharomyces pastorianus  
US-11-217-529-82746

Query Match 72.6%; Score 13.8; DB 7; Length 696;  
Best Local Similarity 88.2%; Pred. No. 2.4e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TTTACTTCATAGTCTTT 18  
||| |||||||  
Db 602 TTAATTCATAGTCTTT 586

## RESULT 52

US-11-217-529-77789/c  
; Sequence 77789, Application US/11217529  
; Publication No. US20060099612A1  
; GENERAL INFORMATION:  
; APPLICANT: SUNTORY LIMITED  
; APPLICANT: NAKAO, YOSHIHIRO  
; APPLICANT: NAKAMURA, NORIHISA  
; APPLICANT: KODAMA, YUKIKO  
; APPLICANT: FUJIMURA, TOMOKO  
; APPLICANT: ASHIKARI, TOSHIHIKO  
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS  
; FILE REFERENCE: S-38-285  
; CURRENT APPLICATION NUMBER: US/11/217,529  
; CURRENT FILING DATE: 2005-09-02  
; PRIOR FILING DATE: 2004-09-02  
; NUMBER OF SEQ ID NOS: 197023  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 77789  
; LENGTH: 729  
; TYPE: DNA  
; ORGANISM: Saccharomyces pastorianus  
US-11-217-529-77789

Query Match 72.6%; Score 13.8; DB 7; Length 729;  
Best Local Similarity 88.2%; Pred. No. 2.4e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 TTACTTCATAGTCTTTG 19  
||| |||||||  
Db 454 TTATCATAGTCTTTG 438

## RESULT 53

US-10-953-349-8745  
; Sequence 8745, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; TITLE OF INVENTION: ENCODED THERBY  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252

```

; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 8745
; LENGTH: 841
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-10-953-349-8745

```

Query Match	72.6%	Score 13.8;	DB 6;	Length 841;
Best Local Similarity	89.2%;	Pred. No. 2.5e+02;		
Matches 15;	Conservative	0;	Mismatches 2;	Indels 0;
Gaps 0;				

Qy 1 CTTTACTTCATAGTCTT 17  
|||  
Db 47 CTTTACTTTATAGTTT 63

```

RESULT 54
US-11-217-529-166678/c
; Sequence 166678, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHIISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 166678
; LENGTH: 867
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-166678

```

Query Match	72.6%	Score 13.8	DB 7	Length 867
Best Local Similarity	88.2%	Pred. NO. 2.5e+02		
Matches 15	Conservative	0	Mismatches 2	Indels 0
Matches 15	Conservative	0	Mismatches 2	Gaps 0

Qy 2 TTTACTTCATAGTCTTT 18  
|||  
db 845 TCTACTTCATAGCCTTT 829

```

RESULT 55
US-10-471-571A-3451/c
; Sequence 3451, Application US/10471571A
; Publication No. US20060115490A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026927WO
; CURRENT APPLICATION NUMBER: US/10/471,571A
; CURRENT FILING DATE: 2003-09-12
; PRIOR APPLICATION NUMBER: GB-0107661.1
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 5642
; SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 3451
; LENGTH: 885
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-10-471-571A-3451

```

Query Match	72.6%	Score 13.8;	DB 6;	Length 885;
Best Local Similarity	88.2%;	Pred. No. 2.5e+02;		
Matches 15:	Conservative	0;	Mismatches 2;	Indels 0;
	0;			Gaps 0;

Qy            3 TTACTTCATAGTCTTG 19  
             ||| ||| ||| ||| |||  
Db            793 TTAGTTTCATAGTCTTG 777

```

RESULT 56
US-10-471-571A-3035/c
; Sequence 3035, Application US/10471571A
; Publication No. US20060115490A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026927WO
; CURRENT APPLICATION NUMBER: US/10/471,571A
; CURRENT FILING DATE: 2003-09-12
; PRIOR APPLICATION NUMBER: GB-0107661.1
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 5642
; SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 3035
; LENGTH: 900
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-10-471-571A-3035

```

Query Match	72.6%	Score 13.8;	DB 6;	Length 900;
Best Local Similarity	88.2%	Pred. No. 2.5e+02;		
Matches 15:	Conservative	0;	Mismatches 2;	Indels 0;
	Gaps	0;		

Qy 2 TTTACTTCATAGTCTTT 18  
Db 626 TTTACTTCATTTTCTTT 610

```

RESULT 57
US-10-953-349-7110
; Sequence 7110, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nickolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE OF INVENTION: ENCODED THEREBY
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 7110
; LENGTH: 946
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-10-953-349-7110

```

Query Match	Score	DB 6;	Length
Best Local Similarity	85.2%;	Pred. No. 2.5e+02;	
Matches 15: Conservative	0: Mismatches 2;	Indels 0;	Gaps 0;
	72.6%;	Score 13.8;	DB 6; Length 946;

Qy 1 CTTTACTTCATAGTCTT 17  
db 93 CTCTACTTCATATTCTT 109

RESULT 58  
US-10-471-571A-3891/c  
; Sequence 3891, Application US/10471571A  
; Publication No. US20060115490A1  
; GENERAL INFORMATION:  
; APPLICANT: CHIRON SpA  
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS PROTEINS AND NUCLEIC ACIDS  
; FILE REFERENCE: P026927WO  
; CURRENT APPLICATION NUMBER: US/10/471,571A  
; CURRENT FILING DATE: 2003-09-12



Matches 15; conservative 0; mismatches 2; indels 0; gaps 0;

QY 2 TTTACTTCATAGTCTTT 18  
|||||  
Db 22 TTTACTTCAGAGTCTGT 38  
|||||

## RESULT 63

US-10-953-349-20194  
; Sequence 20194, Application US/109533349  
; Publication No. US20060107345A1

## GENERAL INFORMATION:

; APPLICANT: ALEXANDROV, Nikolai et al.

; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES

; FILE OF INVENTION: ENCODED THERBY

; FILE REFERENCE: 2750-1579PUS2

; CURRENT APPLICATION NUMBER: US/10/953,349

; CURRENT FILING DATE: 2004-09-30

; NUMBER OF SEQ ID NOS: 40252

; SOFTWARE: PatentIn version 3.3

; SEQ ID NO 20194

; LENGTH: 1240

; TYPE: DNA

; ORGANISM: Glycine max

US-10-953-349-20194

Query Match 72.6%; Score 13.8; DB 6; Length 1240;

Best Local Similarity 88.2%; Pred. No. 2.6e+02;

Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCTTT 17  
|||||  
Db 592 CTTTACATCATATCTTT 608  
|||||

## RESULT 64

US-11-217-529-78180/c  
; Sequence 78180, Application US/11217529  
; Publication No. US2006099612A1

## GENERAL INFORMATION:

; APPLICANT: SUNTORY LIMITED

; APPLICANT: NAKAO, YOSHIIRO

; APPLICANT: NAKAMURA, NORIHISA

; APPLICANT: KODAMA, YUKIKO

; APPLICANT: FUJIMURA, TOMOKO

; APPLICANT: ASHIKARI, TOSHIHIKO

; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS

; FILE REFERENCE: S-38-285

; CURRENT APPLICATION NUMBER: US/11/217,529

; CURRENT FILING DATE: 2005-09-02

; PRIOR APPLICATION NUMBER: US 10/932,182

; PRIOR FILING DATE: 2004-09-02

; NUMBER OF SEQ ID NOS: 197023

; SOFTWARE: PatentIn version 3.3

; SEQ ID NO 78180

; LENGTH: 1245

; TYPE: DNA

; ORGANISM: Saccharomyces pastorianus

US-11-217-529-78180

Query Match 72.6%; Score 13.8; DB 7; Length 1245;

Best Local Similarity 88.2%; Pred. No. 2.6e+02;

Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCTTT 17  
|||||  
Db 132 CTTTACTTCATCTCTTT 116  
|||||

## RESULT 65

US-10-953-349-1160

; Sequence 1160, Application US/109533349

; Publication No. US20060107345A1

## GENERAL INFORMATION:

; APPLICANT: ALEXANDROV, Nikolai et al.

; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES

; FILE OF INVENTION: ENCODED THERBY

; FILE REFERENCE: 2750-1579PUS2

; CURRENT APPLICATION NUMBER: US/10/953,349

; CURRENT FILING DATE: 2004-09-30

; NUMBER OF SEQ ID NOS: 40252

; SOFTWARE: PatentIn version 3.3

; SEQ ID NO 1160

; LENGTH: 1272

; TYPE: DNA

; ORGANISM: Arabidopsis thaliana

US-10-953-349-1160

Query Match 72.6%; Score 13.8; DB 6; Length 1272;

Best Local Similarity 88.2%; Pred. No. 2.6e+02;

Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 TTACTTCATAGTCTTTG 19  
|||||  
Db 1106 TTACTTCCTAGTCCTTG 1122  
|||||

## RESULT 66

US-10-953-349-20559/c

; Sequence 20559, Application US/109533349

; Publication No. US20060107345A1

## GENERAL INFORMATION:

; APPLICANT: ALEXANDROV, Nikolai et al.

; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES

; FILE OF INVENTION: ENCODED THERBY

; FILE REFERENCE: 2750-1579PUS2

; CURRENT APPLICATION NUMBER: US/10/953,349

; CURRENT FILING DATE: 2004-09-30

; NUMBER OF SEQ ID NOS: 40252

; SOFTWARE: PatentIn version 3.3

; SEQ ID NO 20559

; LENGTH: 1315

; TYPE: DNA

; ORGANISM: Glycine max

US-10-953-349-20559

Query Match 72.6%; Score 13.8; DB 6; Length 1315;

Best Local Similarity 88.2%; Pred. No. 2.6e+02;

Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTTTACTTCATAGTCTTT 17  
|||||  
Db 1155 CTTTACTTCATGTATT 1139  
|||||

## RESULT 67

US-10-953-349-15758/c

; Sequence 15758, Application US/109533349

; Publication No. US20060107345A1

## GENERAL INFORMATION:

; APPLICANT: ALEXANDROV, Nikolai et al.

; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES

; FILE OF INVENTION: ENCODED THERBY

; FILE REFERENCE: 2750-1579PUS2

; CURRENT APPLICATION NUMBER: US/10/953,349

; CURRENT FILING DATE: 2004-09-30

; NUMBER OF SEQ ID NOS: 40252

; SOFTWARE: PatentIn version 3.3

; SEQ ID NO 15758

; LENGTH: 1332

; TYPE: DNA

; ORGANISM: Glycine max

US-10-953-349-15758

Query Match 72.6%; Score 13.8; DB 6; Length 1332;

Best Local Similarity 88.2%; Pred. No. 2.6e+02;

Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy 2 TTTACTTCATAGTCCTT 18
    ||||| ||| |||||
Db 708 TTTACTTAATATCTTT 692

RESULT 68
US-10-953-349-17823
; Sequence 17823, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 17823
; LENGTH: 1333
; TYPE: DNA
; ORGANISM: Glycine max
US-10-953-349-17823

Query Match 72.6%; Score 13.8; DB 6; Length 1333;
Best Local Similarity 88.2%; Pred. No. 2.6e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TTTACTTCATAGTCCTT 18
    ||||| ||| |||||
Db 630 TATACCTCACAGTCCTT 646

RESULT 69
US-10-471-571A-4559
; Sequence 4559, Application US/10471571A
; Publication No. US20060115490A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026927WO
; CURRENT APPLICATION NUMBER: US/10/471,571A
; CURRENT FILING DATE: 2003-09-12
; PRIOR APPLICATION NUMBER: GB-0107661.1
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 5642
; SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 4559
; LENGTH: 1341
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-10-471-571A-4559

Query Match 72.6%; Score 13.8; DB 6; Length 1341;
Best Local Similarity 88.2%; Pred. No. 2.6e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TTTACTTCATAGTCCTT 18
    ||||| ||| |||||
Db 371 TCTACTTCATAGTCGTT 387

RESULT 70
US-10-953-349-24191
; Sequence 24191, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
```

```
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 24191
; LENGTH: 1404
; TYPE: DNA
; ORGANISM: Glycine max
US-10-953-349-24191

Query Match 72.6%; Score 13.8; DB 6; Length 1404;
Best Local Similarity 88.2%; Pred. No. 2.6e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TTTACTTCATAGTCCTT 18
    ||||| ||| |||||
Db 846 TATACCTCACAGTCCTT 862

RESULT 71
US-10-953-349-7778/c
; Sequence 7778, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 7778
; LENGTH: 1420
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-10-953-349-7778

Query Match 72.6%; Score 13.8; DB 6; Length 1420;
Best Local Similarity 88.2%; Pred. No. 2.6e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCCTT 17
    ||||| ||| |||||
Db 669 CTTTAATTCAAAGTCCTT 653

RESULT 72
US-11-217-529-82109/c
; Sequence 82109, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 82109
; LENGTH: 1461
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-82109

Query Match 72.6%; Score 13.8; DB 7; Length 1461;
Best Local Similarity 88.2%; Pred. No. 2.6e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      2 TTTACTTCATAGTCTTT 18
      ||||| ||||| |||||
Db      1415 TTTACATCATCGTCTTT 1399

RESULT 73
US-11-217-529-5036/c
; Sequence 5036, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 5036
; LENGTH: 1485
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-5036

Query Match      72.6%; Score 13.8; DB 7; Length 1485;
Best Local Similarity 88.2%; Pred. No. 2.6e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 TTTACTTCATAGTCTTT 18
      ||||| ||||| |||||
Db      1424 TCTACTTCATAGTCTGT 1408

RESULT 74
US-11-217-529-82729/c
; Sequence 82729, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 82729
; LENGTH: 1605
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-82729

Query Match      72.6%; Score 13.8; DB 7; Length 1605;
Best Local Similarity 88.2%; Pred. No. 2.7e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 TTTACTTCATAGTCTTT 18
      ||||| ||||| |||||
Db      692 TTTATTTCATGTCTTT 676
```

```
RESULT 75
US-11-217-529-421/c
; Sequence 421, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 421
; LENGTH: 1704
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-421

Query Match      72.6%; Score 13.8; DB 7; Length 1704;
Best Local Similarity 88.2%; Pred. No. 2.7e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 TTTACTTCATAGTCTTT 18
      ||||| ||||| |||||
Db      1107 TTCACTCCATAGTCTTT 1091

RESULT 76
US-11-217-529-4484/c
; Sequence 4484, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 4484
; LENGTH: 1800
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-4484

Query Match      72.6%; Score 13.8; DB 7; Length 1800;
Best Local Similarity 88.2%; Pred. No. 2.7e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 TTTACTTCATAGTCTTT 18
      ||||| ||||| |||||
Db      166 TTTCTTCATAGTATTT 150

RESULT 77
US-10-953-349-20342
; Sequence 20342, Application US/10953349
; Publication No. US20060107345A1
```

```
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE OF INVENTION: ENCODED THERY
; FILE REFERENCE: 2750-1579FUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 20342
; LENGTH: 1806
; TYPE: DNA
; ORGANISM: Glycine max
US-10-953-349-20342

Query Match          72.6%; Score 13.8; DB 6; Length 1806;
Best Local Similarity 88.2%; Pred. No. 2.7e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TTACTTCATAGTCTTT 18
   ||||| ||||| |||||
Db 1538 TATACCTCACAGTCTTT 1554

RESULT 78
US-10-953-349-1741
; Sequence 1741, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE OF INVENTION: ENCODED THERY
; FILE REFERENCE: 2750-1579FUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1741
; LENGTH: 1843
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-10-953-349-1741

Query Match          72.6%; Score 13.8; DB 6; Length 1843;
Best Local Similarity 88.2%; Pred. No. 2.7e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 TTACTTCATAGTCTTTG 19
   ||||| ||||| |||||
Db 1486 TTACTTGATATCTTTG 1502

RESULT 79
US-11-217-529-534/c
; Sequence 534, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 534
; LENGTH: 1854
```

```
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-534

Query Match          72.6%; Score 13.8; DB 7; Length 1854;
Best Local Similarity 88.2%; Pred. No. 2.7e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TTACTTCATAGTCTTT 18
   ||||| ||||| |||||
Db 1607 TTACTTCATTATCTTT 1591

RESULT 80
US-10-953-349-18787
; Sequence 18787, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE OF INVENTION: ENCODED THERY
; FILE REFERENCE: 2750-1579FUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 18787
; LENGTH: 1861
; TYPE: DNA
; ORGANISM: Glycine max
US-10-953-349-18787

Query Match          72.6%; Score 13.8; DB 6; Length 1861;
Best Local Similarity 88.2%; Pred. No. 2.7e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTTT 17
   ||||| ||||| |||||
Db 109 CTTTCTTCATGTCTTT 125

RESULT 81
US-10-471-571A-4139/c
; Sequence 4139, Application US/10471571A
; Publication No. US20060115490A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON Spa
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026927W0
; CURRENT APPLICATION NUMBER: US/10/471,571A
; CURRENT FILING DATE: 2003-09-12
; PRIOR APPLICATION NUMBER: GB-0107661.1
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 5642
; SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 4139
; LENGTH: 1899
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-10-471-571A-4139

Query Match          72.6%; Score 13.8; DB 6; Length 1899;
Best Local Similarity 88.2%; Pred. No. 2.7e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 TTACTTCATAGTCTTTG 19
   ||||| ||||| |||||
Db 1378 TTTCTTGATAGTCTTTG 1362

RESULT 82
US-11-217-529-75539/c
; Sequence 75539, Application US/11217529
```

```
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 75539
; LENGTH: 2004
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-75539

Query Match      72.6%; Score 13.8; DB 7; Length 2004;
Best Local Similarity 88.2%; Pred. No. 2.7e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      3 TTACTTCATAGTCTTTG 19
Db      511 TTCTTCATATCTTTG 495

RESULT 83
US-11-217-529-565/c
; Sequence 565, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 565
; LENGTH: 2031
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-565

Query Match      72.6%; Score 13.8; DB 7; Length 2031;
Best Local Similarity 88.2%; Pred. No. 2.7e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      3 TTACTTCATAGTCTTTG 19
Db      997 TTCTTCATATCTTTG 981

RESULT 84
US-10-471-571A-1409/c
; Sequence 1409, Application US/10471571A
; Publication No. US20060115490A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026927WO
```

```
; CURRENT APPLICATION NUMBER: US/10/471,571A
; CURRENT FILING DATE: 2003-09-12
; PRIOR APPLICATION NUMBER: GB-0107661.1
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 5642
; SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 1409
; LENGTH: 2190
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-10-471-571A-1409

Query Match      72.6%; Score 13.8; DB 6; Length 2190;
Best Local Similarity 88.2%; Pred. No. 2.7e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      2 TTTTACTTCATAGTCTTT 18
Db      1559 TTTTCTTCATAGTCTTT 1543

RESULT 85
US-10-953-349-8153
; Sequence 8153, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 8153
; LENGTH: 2320
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-10-953-349-8153

Query Match      72.6%; Score 13.8; DB 6; Length 2320;
Best Local Similarity 88.2%; Pred. No. 2.8e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 CTTTACTTCATAGTCTT 17
Db      102 CTTTACTTAATCGTCTT 118

RESULT 86
US-11-293-697-52/c
; Sequence 52, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: HI-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 52
; LENGTH: 2577
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-293-697-52

Query Match      72.6%; Score 13.8; DB 7; Length 2577;
Best Local Similarity 88.2%; Pred. No. 2.8e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

Qy 2 TTTACTTCATAGTCTTT 18  
||||| ||||| |||||  
Db 1098 TTTACATCATAGTCTAT 1082

RESULT 87  
US-11-293-697-749  
; Sequence 749, Application US/11293697  
; Publication No. US20060105376A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: Novel full length cDNA  
; FILE REFERENCE: HI-A0106  
; CURRENT APPLICATION NUMBER: US/11/293,697  
; CURRENT FILING DATE: 2005-12-05  
; PRIOR APPLICATION NUMBER: US/10/108,260  
; PRIOR FILING DATE: 2002-03-28  
; NUMBER OF SEQ ID NOS: 5458  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 749  
; LENGTH: 2650  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-293-697-749

Query Match 72.6%; Score 13.8; DB 7; Length 2650;  
Best Local Similarity 88.2%; Pred. No. 2.8e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TTTACTTCATAGTCTTT 18  
||||| ||||| |||||  
Db 742 TTTACTTCATCATCTTT 758

RESULT 88  
US-11-293-697-2321/c  
; Sequence 2321, Application US/11293697  
; Publication No. US20060105376A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: Novel full length cDNA  
; FILE REFERENCE: HI-A0106  
; CURRENT APPLICATION NUMBER: US/11/293,697  
; CURRENT FILING DATE: 2005-12-05  
; PRIOR APPLICATION NUMBER: US/10/108,260  
; PRIOR FILING DATE: 2002-03-28  
; NUMBER OF SEQ ID NOS: 5458  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2321  
; LENGTH: 2854  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-293-697-2321

Query Match 72.6%; Score 13.8; DB 7; Length 2854;  
Best Local Similarity 88.2%; Pred. No. 2.8e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TTTACTTCATAGTCTTT 18  
||||| ||||| |||||  
Db 2397 TTTACTTCAAAGTCATT 2381

RESULT 89  
US-11-217-529-82302/c  
; Sequence 82302, Application US/11217529  
; Publication No. US20060099612A1  
; GENERAL INFORMATION:  
; APPLICANT: SUNTORY LIMITED  
; APPLICANT: NAKAO, YOSHIHIRO  
; APPLICANT: NAKAMURA, NORIHISA  
; APPLICANT: KODAMA, YUKIKO  
; APPLICANT: FUJIMURA, TOMOKO

; APPLICANT: ASHIKARI, TOSHIHIKO  
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS  
; FILE REFERENCE: S-38-285  
; CURRENT APPLICATION NUMBER: US/11/217,529  
; CURRENT FILING DATE: 2005-09-02  
; PRIOR APPLICATION NUMBER: US 10/932,182  
; PRIOR FILING DATE: 2004-09-02  
; NUMBER OF SEQ ID NOS: 197023  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 82302  
; LENGTH: 3102  
; TYPE: DNA  
; ORGANISM: Saccharomyces pastorianus  
US-11-217-529-82302

Query Match 72.6%; Score 13.8; DB 7; Length 3102;  
Best Local Similarity 88.2%; Pred. No. 2.9e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTTTACTTCATAGTCTT 17  
||||| ||||| |||||  
Db 2100 CTTTTCATAGTCTT 2084

RESULT 90  
US-11-217-529-4879/c  
; Sequence 4879, Application US/11217529  
; Publication No. US20060099612A1  
; GENERAL INFORMATION:  
; APPLICANT: SUNTORY LIMITED  
; APPLICANT: NAKAO, YOSHIHIRO  
; APPLICANT: NAKAMURA, NORIHISA  
; APPLICANT: KODAMA, YUKIKO  
; APPLICANT: FUJIMURA, TOMOKO  
; APPLICANT: ASHIKARI, TOSHIHIKO  
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS  
; FILE REFERENCE: S-38-285  
; CURRENT APPLICATION NUMBER: US/11/217,529  
; CURRENT FILING DATE: 2005-09-02  
; PRIOR APPLICATION NUMBER: US 10/932,182  
; PRIOR FILING DATE: 2004-09-02  
; NUMBER OF SEQ ID NOS: 197023  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 4879  
; LENGTH: 3273  
; TYPE: DNA  
; ORGANISM: Saccharomyces pastorianus  
US-11-217-529-4879

Query Match 72.6%; Score 13.8; DB 7; Length 3273;  
Best Local Similarity 88.2%; Pred. No. 2.9e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TTTACTTCATAGTCTTT 18  
||||| ||||| |||||  
Db 845 TCTACTTCATAGCCTTT 829

RESULT 91  
US-11-217-529-77385/c  
; Sequence 77385, Application US/11217529  
; Publication No. US20060099612A1  
; GENERAL INFORMATION:  
; APPLICANT: SUNTORY LIMITED  
; APPLICANT: NAKAO, YOSHIHIRO  
; APPLICANT: NAKAMURA, NORIHISA  
; APPLICANT: KODAMA, YUKIKO  
; APPLICANT: FUJIMURA, TOMOKO  
; APPLICANT: ASHIKARI, TOSHIHIKO  
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS  
; FILE REFERENCE: S-38-285  
; CURRENT APPLICATION NUMBER: US/11/217,529  
; CURRENT FILING DATE: 2005-09-02

```
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 77385
; LENGTH: 3447
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (2669)..(2669)
; OTHER INFORMATION: a, c, g, t, unknown, or other
US-11-217-529-77385

Query Match      72.6%; Score 13.8; DB 7; Length 3447;
Best Local Similarity 88.2%; Pred. No. 2.9e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 TTACTTCATAGTCCTT 18
Db      1123 TTAACCTCATAGTCCTT 1107

RESULT 92
US-11-217-529-3930
; Sequence 3930, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 3930
; LENGTH: 3528
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-3930

Query Match      72.6%; Score 13.8; DB 7; Length 3528;
Best Local Similarity 88.2%; Pred. No. 2.9e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      3 TTACTTCATAGTCCTTGG 19
Db      1776 TTCCTTTATAGTCCTTGG 1792

RESULT 93
US-10-953-349-2088/c
; Sequence 2088, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 2088
; LENGTH: 4746
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-10-953-349-2088

Query Match      72.6%; Score 13.8; DB 6; Length 4746;
Best Local Similarity 88.2%; Pred. No. 3e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 CTTTACTTCATAGTCCTT 17
Db      2284 CTTTACATCAGAGTCCTT 2268

RESULT 94
US-11-217-529-5476/c
; Sequence 5476, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 5476
; LENGTH: 5283
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-5476

Query Match      72.6%; Score 13.8; DB 7; Length 5283;
Best Local Similarity 88.2%; Pred. No. 3e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 TTTTACTTCATAGTCCTT 18
Db      2918 TTTTCTTCATATCTT 2902

RESULT 95
US-10-506-549-3
; Sequence 3, Application US/10506549
; Publication No. US20060100417A1
; GENERAL INFORMATION:
; APPLICANT: APPLERA CORPORATION
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL001361-US
; CURRENT APPLICATION NUMBER: US/10/506,549
; CURRENT FILING DATE: 2004-09-03
; PRIOR APPLICATION NUMBER: 60/361,343
; PRIOR FILING DATE: 2002-03-05
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 394191
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(394191)
; OTHER INFORMATION: n = A,T,C or G
US-10-506-549-3

Query Match      72.6%; Score 13.8; DB 6; Length 394191;
```



Best Local Similarity 88.2%; Pred. No. 2.5e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TTTTACTTCATAGTCTTT 18  
Db 361068 TTGACTTCAAAGTCTTT 361084

## RESULT 96

US-11-217-529-74966  
; Sequence 74966, Application US/11217529  
; Publication No. US20060099612A1  
; GENERAL INFORMATION:  
; APPLICANT: SUNTORY LIMITED  
; APPLICANT: NAKAO, YOSHIHIRO  
; APPLICANT: NAKAMURA, NORIHISA  
; APPLICANT: KODAMA, YUKIKO  
; APPLICANT: FUJIMURA, TOMOKO  
; APPLICANT: ASHIKARI, TOSHIHIKO  
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS  
; FILE REFERENCE: S-38-285  
; CURRENT APPLICATION NUMBER: US/11/217,529  
; CURRENT FILING DATE: 2005-09-02  
; PRIOR APPLICATION NUMBER: US 10/932,182  
; PRIOR FILING DATE: 2004-09-02  
; NUMBER OF SEQ ID NOS: 197023  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 74966  
; LENGTH: 25  
; TYPE: DNA  
; ORGANISM: Saccharomyces pastorianus  
US-11-217-529-74966

Query Match 70.5%; Score 13.4; DB 7; Length 25;  
Best Local Similarity 93.3%; Pred. No. 2.6e+02;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TTTTACTTCATAGTCT 16  
Db 3 TTTTACTTCAGTCT 17

## RESULT 97

US-10-471-571A-1517/c  
; Sequence 1517, Application US/10471571A  
; Publication No. US20060115490A1  
; GENERAL INFORMATION:  
; APPLICANT: CHIRON SpA  
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS PROTEINS AND NUCLEIC ACIDS  
; FILE REFERENCE: P026927W0  
; CURRENT APPLICATION NUMBER: US/10/471,571A  
; CURRENT FILING DATE: 2003-09-12  
; PRIOR APPLICATION NUMBER: GB-0107661.1  
; PRIOR FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 5642  
; SOFTWARE: SeqWin99, version 1.03  
; SEQ ID NO 1517  
; LENGTH: 552  
; TYPE: DNA  
; ORGANISM: Staphylococcus aureus  
US-10-471-571A-1517

Query Match 70.5%; Score 13.4; DB 6; Length 552;  
Best Local Similarity 93.3%; Pred. No. 3.6e+02;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 TACTTCATAGTCTTT 18  
Db 408 TACTTCATGCTCTTT 394

## RESULT 98

US-10-953-349-15823/c

; Sequence 15823, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nickolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 15823  
; LENGTH: 878  
; TYPE: DNA  
; ORGANISM: Glycine max  
US-10-953-349-15823

Query Match 70.5%; Score 13.4; DB 6; Length 878;  
Best Local Similarity 93.3%; Pred. No. 3.8e+02;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 TACTTCATAGTCTTT 18  
Db 336 TACTTCATGCTCTTT 322

## RESULT 99

US-11-217-529-269/c  
; Sequence 269, Application US/11217529  
; Publication No. US20060099612A1  
; GENERAL INFORMATION:  
; APPLICANT: SUNTORY LIMITED  
; APPLICANT: NAKAO, YOSHIHIRO  
; APPLICANT: NAKAMURA, NORIHISA  
; APPLICANT: KODAMA, YUKIKO  
; APPLICANT: FUJIMURA, TOMOKO  
; APPLICANT: ASHIKARI, TOSHIHIKO  
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS  
; FILE REFERENCE: S-38-285  
; CURRENT APPLICATION NUMBER: US/11/217,529  
; CURRENT FILING DATE: 2005-09-02  
; PRIOR APPLICATION NUMBER: US 10/932,182  
; PRIOR FILING DATE: 2004-09-02  
; NUMBER OF SEQ ID NOS: 197023  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 269  
; LENGTH: 966  
; TYPE: DNA  
; ORGANISM: Saccharomyces pastorianus  
US-11-217-529-269

Query Match 70.5%; Score 13.4; DB 7; Length 966;  
Best Local Similarity 93.3%; Pred. No. 3.9e+02;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 5 ACTTCATAGTCTTTG 19  
Db 863 ACTTCATGCTCTTG 849

## RESULT 100

US-11-253-300-6  
; Sequence 6, Application US/11253300  
; Publication No. US20060107349A1  
; GENERAL INFORMATION:  
; APPLICANT: ZHANG, JIYI  
; APPLICANT: WANG, ZENGJU  
; TITLE OF INVENTION: INCREASED WAX PRODUCTION IN PLANTS  
; FILE REFERENCE: NELE:045US  
; CURRENT APPLICATION NUMBER: US/11/253,300  
; CURRENT FILING DATE: 2005-10-18  
; PRIOR APPLICATION NUMBER: 60/619,801  
; PRIOR FILING DATE: 2004-10-18

•  
; NUMBER OF SEQ ID NOS: 46  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 1023  
; TYPE: DNA  
; ORGANISM: Medicago truncatula  
US-11-253-300-6

Query Match 70.5%; Score 13.4; DB 7; Length 1023;  
Best Local Similarity 93.3%; Pred. No. 3.9e+02;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 ACTTCATAGCTTTG 19  
||| ||||| |||||  
Db 200 ACTTCATAGGCTTTG 214

Search completed: June 10, 2006, 15:37:40  
Job time : 17.1833 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2006, 15:18:44 ; Search time 1145.2 Seconds  
(without alignments)  
2345.260 Million cell updates/sec

Title: US-10-600-816-33

Perfect score: 33.2

Sequence: 1 gcncaaygntggcctcnc.....ayaargaytgargttnaar 42

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 6366136 seqs, 31973710525 residues

Total number of hits satisfying chosen parameters: 12732272

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 1000 summaries

Database :

GenEmbl.\*

1: gb\_env.\*

2: gb\_pat.\*

3: gb\_ph.\*

4: gb\_pl.\*

5: gb\_pr.\*

6: gb\_ro.\*

7: gb\_sts.\*

8: gb\_sy.\*

9: gb\_un.\*

10: gb\_vi.\*

11: gb\_ov.\*

12: gb\_htg.\*

13: gb\_in.\*

14: gb\_om.\*

15: gb\_ba.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	30	90.4	603	2	BD233463
2	30	90.4	1212	2	BD209699
3	30	90.4	1212	2	AK341505
4	30	90.4	1619	2	AX078375
5	30	90.4	1718	2	BD233473
6	30	90.4	1826	5	AK172760
7	30	90.4	2290	2	CO723177
8	30	90.4	2296	5	BC003665
9	30	90.4	2297	5	AF506289
10	30	90.4	2302	2	AX930411
11	30	90.4	2302	5	AF095448
12	30	90.4	2446	2	BD156680
13	30	90.4	2446	2	AX877483
14	30	90.4	2446	5	AK001761
15	30	90.4	2456	2	CO981495
16	30	90.4	2456	2	DD210040
17	30	90.4	2456	2	AX549168
18	30	90.4	3057	5	AK123672

19	30	90.4	3371	2	AX188348
20	30	90.4	6730	2	CO894732
c 21	30	90.4	161577	5	AC007688
22	27.4	82.5	680	2	BD150346
23	27.4	82.5	680	2	AX870284
24	22.2	66.9	224653	12	AC149246
c 25	22	66.3	179391	5	AC012558
26	22	66.3	202050	12	AC060229
c 27	22	66.3	222810	5	AC018977
28	21.6	65.1	179591	5	AC093902
c 29	21.6	65.1	182332	6	AL606745
c 30	21.6	65.1	203317	12	AC021062
31	21.4	64.5	859	4	AK111295
32	21.4	64.5	21062	13	CEVY10G11R
33	21.4	64.5	110000	4	AP008217_262
34	21.4	64.5	133581	4	AC134924
35	21.4	64.5	164168	5	AL135927
c 36	21.4	64.5	164179	5	AC007227
37	21.4	64.5	186979	12	AC158014
38	21.4	64.5	196244	12	AC152284
39	21.2	63.9	1604	15	AF110497
40	21.2	63.9	110000	15	EX897700_01
41	21.2	63.9	110000	15	EX897700_02
42	21.2	63.9	203946	12	AC069465
c 43	21.2	63.9	235241	6	AL928893
c 44	21	63.3	98142	12	AC159685
45	21	63.3	142190	5	AC066598
46	21	63.3	151830	12	AC024146
c 47	21	63.3	152580	5	AC024164
48	21	63.3	154010	12	AC024219
c 49	21	63.3	157590	12	AC172585
50	21	63.3	165646	12	CT027633
51	21	63.3	168908	11	CR361547
c 52	21	63.3	178752	5	AC021889
c 53	21	63.3	181736	5	AC087589
54	21	63.3	185496	5	AC018832
c 55	21	63.3	191265	12	AC068996
56	21	63.3	197755	12	CT573362
57	21	63.3	215523	12	CT172772
c 58	21	63.3	228141	6	AC154912
59	21	63.3	229726	12	AC095735
c 60	21	63.3	235642	6	CT010441
61	21	63.3	246535	12	AC109060
c 62	20.8	62.7	571	7	BV053582
63	20.8	62.7	3378	11	BC078095
64	20.8	62.7	81744	11	EX927176
c 65	20.8	62.7	110000	15	AE014295_15
66	20.8	62.7	139509	12	CR354544
c 67	20.8	62.7	154959	6	AL662881
c 68	20.8	62.7	165987	11	AC147883
c 69	20.8	62.7	171978	6	AC131779
c 70	20.8	62.7	174929	12	AC162371
71	20.8	62.7	194709	12	AC132640
72	20.8	62.7	195892	12	AC126211
c 73	20.8	62.7	210916	12	CT573319
c 74	20.8	62.7	212324	12	AC096459
c 75	20.8	62.7	219429	12	AC137342
76	20.8	62.7	232234	12	AC108634
c 77	20.8	62.7	240216	12	AC107096
78	20.8	62.7	241557	12	AC098000
79	20.8	62.7	243005	12	AC134740
80	20.8	62.7	287291	12	AC115384
81	20.8	62.7	319548	12	AC110863
c 82	20.8	62.7	349980	2	AX492783
c 83	20.8	62.7	349980	2	AX553950
84	20.6	62.0	770	7	BV639777
85	20.6	62.0	9060	15	MYCPM358
86	20.6	62.0	9720	15	MYCPM358
87	20.6	62.0	44308	12	AC100253
c 88	20.6	62.0	84234	5	AE017975
89	20.6	62.0	110000	15	AE015450_04
c 90	20.6	62.0	112585	12	AM050691
c 91	20.6	62.0	142445	12	AC146668

C 92	20.6	62.0	154471	5	AC022523	AC022523 Homo sapi	165	20.2	60.8	209476	12	AC021027	AC021027 Homo sapi
C 93	20.6	62.0	165165	12	CT025647	CT025647 Danio rer	C 166	20.2	60.8	213756	12	AC164200	AC164200 Bos tauru
C 94	20.6	62.0	171116	12	AP003651	AP003651 Mus muscu	C 167	20.2	60.8	214870	6	AL606496	AL606496 Mouse DNA
C 95	20.6	62.0	174250	6	AC164162	AC164162 Mus muscu	C 168	20.2	60.8	215467	12	AC013420	AC013420 Homo sapi
C 96	20.6	62.0	186376	6	CT027693	CT027693 Mouse DNA	C 169	20.2	60.8	219854	11	AC140791	AC140791 Gallus ga
C 97	20.6	62.0	191693	6	AC123073	AC123073 Mus muscu	C 170	20.2	60.8	228115	12	AC128376	AC128376 Rattus no
C 98	20.6	62.0	193668	6	AC147235	AC147235 Mus muscu	C 171	20.2	60.8	230020	12	AC106534	AC106534 Rattus no
C 99	20.6	62.0	207858	12	AC127132	AC127132 Rattus no	C 172	20.2	60.8	233440	12	AC094560	AC094560 Rattus no
C 100	20.6	62.0	207858	12	AC134718	AC134718 Rattus no	C 173	20.2	60.8	233440	12	AC094560	AC094560 Rattus no
C 101	20.6	62.0	219121	12	AC134718	AC134718 Rattus no	C 174	20.2	60.8	239385	12	AC109686	AC109686 Rattus no
C 102	20.6	62.0	221744	12	AC172135	AC172135 Bos tauru	C 175	20.2	60.8	248669	12	AC127097	AC127097 Rattus no
C 103	20.6	62.0	228501	6	AC094784	AC094784 Rattus no	C 176	20.2	60.8	250600	12	AC136646	AC136646 Rattus no
C 104	20.6	62.0	236068	12	AC121712	AC121712 Rattus no	C 177	20.2	60.8	254885	12	AC097132	AC097132 Rattus no
C 105	20.6	62.0	253387	12	AC133379	AC133379 Rattus no	C 178	20.2	60.8	260290	12	AC097025	AC097025 Rattus no
C 106	20.4	61.4	753	7	BV211953	BV211953 ONMS232 T	C 179	20.2	60.8	262690	12	AC094332	AC094332 Rattus no
C 107	20.4	61.4	1053	2	AR764489	AR764489 Sequence	C 180	20.2	60.8	270023	12	AC094829	AC094829 Rattus no
C 108	20.4	61.4	1074	2	BD162796	BD162796 Novel pol	C 181	20.2	60.8	271603	12	AC106318	AC106318 Rattus no
C 109	20.4	61.4	1201	2	AX120679	AX120679 Sequence	C 182	20.2	60.8	277252	12	AC175278	AC175278 Bos tauru
C 110	20.4	61.4	1857	13	DQ211932	DQ211932 Toxoplasma	C 183	20.2	60.8	349880	2	CQ870195	CQ870195 Sequence
C 111	20.4	61.4	2045	4	SCU15217	UI5217 Saccharomyc	C 184	20.2	60.2	282	7	AB133119	AB133119 Homo sapi
C 112	20.4	61.4	2780	2	AX536762	AX536762 Sequence	C 185	20.2	60.2	669	2	AX349817	AX349817 Sequence
C 113	20.4	61.4	3239	4	SCYOR267C	Z75175 S.cerevisia	C 186	20.2	60.2	718	7	BV611899	BV611899 S21P6209
C 114	20.4	61.4	72325	13	AC025722	AC025722 Caenorhab	C 187	20.2	60.2	2026	4	AK064556	AK064556 Oryza sat
C 115	20.4	61.4	106038	5	AC106747	AC106747 Homo sapi	C 188	20.2	60.2	12548	15	AE002165	AE002165 Chlamydom
C 116	20.4	61.4	110000	4	BA000024_1	Continuation (2 of	C 189	20.2	60.2	12917	5	AL662851	AL662851 Human DNA
C 117	20.4	61.4	110000	4	BA000024_3	Continuation (4 of	C 190	20.2	60.2	68890	5	AC004833	AC004833 Homo sapi
C 118	20.4	61.4	110000	12	AP008211_252	Continuation (253	C 191	20.2	60.2	80181	12	AC162649	AC162649 Bos tauru
C 119	20.4	61.4	110000	12	CR450853_0	CR450853 Danio rer	C 192	20.2	60.2	89319	12	AL929255	AL929255 Homo sapi
C 120	20.4	61.4	110000	15	CR522870_21	Continuation (22 o	C 193	20.2	60.2	107415	5	AL513365	AL513365 Human DNA
C 121	20.4	61.4	110000	15	BA000036_05	Continuation (6 of	C 194	20.2	60.2	110000	2	AR310754_08	Continuation (9 of
C 122	20.4	61.4	120891	12	AC112885	AC011285 Homo sapi	C 195	20.2	60.2	110000	2	AR607478_08	Continuation (9 of
C 123	20.4	61.4	146712	4	AC130604	AC130604 Oryza sat	C 196	20.2	60.2	110000	4	AP008208_064	Continuation (65 o
C 124	20.4	61.4	152945	4	AC130606	AC130606 Oryza sat	C 197	20.2	60.2	110000	12	CT005244_3	Continuation (9 of
C 125	20.4	61.4	157572	5	AL513210	AL513210 Human DNA	C 198	20.2	60.2	110000	15	AE001363_08	Continuation (9 of
C 126	20.4	61.4	161354	12	AC113374	AC113374 Homo sapi	C 199	20.2	60.2	110000	15	BA000008_08	Continuation (9 of
C 127	20.4	61.4	167905	5	AC116606	AC116606 Homo sapi	C 200	20.2	60.2	110752	11	BX649506	Continuation (9 of
C 128	20.4	61.4	174884	11	BX2963542	BX2963542 Zebrafish	C 201	20.2	60.2	118572	12	AC181623	Continuation (9 of
C 129	20.4	61.4	177857	12	AC144030	AC144030 Macaca mu	C 202	20.2	60.2	128729	5	AL133408	Continuation (9 of
C 130	20.4	61.4	187845	12	AC118983	AC118983 Rattus no	C 203	20.2	60.2	131538	4	AP005846	Continuation (9 of
C 131	20.4	61.4	203830	12	AC171610	AC171610 Bos tauru	C 204	20.2	60.2	137445	12	AP004780	Continuation (9 of
C 132	20.4	61.4	219390	12	AC114870	AC114870 Rattus no	C 205	20.2	60.2	150221	5	AC004593	Continuation (9 of
C 133	20.4	61.4	221270	12	AC123310	AC123310 Rattus no	C 206	20.2	60.2	156191	5	AC002748	Continuation (9 of
C 134	20.4	61.4	222741	12	AC128894	AC128894 Rattus no	C 207	20.2	60.2	160250	12	AC174408	Continuation (9 of
C 135	20.4	61.4	225487	6	AC162441	AC162441 Mus muscu	C 208	20.2	60.2	161451	5	AL355989	Continuation (9 of
C 136	20.4	61.4	226899	12	AC123119	AC123119 Rattus no	C 209	20.2	60.2	170106	6	AC166995	Continuation (9 of
C 137	20.4	61.4	235135	11	BX649266	BX649266 Zebrafish	C 210	20.2	60.2	170106	6	AC166995	Continuation (9 of
C 138	20.4	61.4	237432	12	AC123182	AC123182 Rattus no	C 211	20.2	60.2	170795	12	AC021346	Continuation (9 of
C 139	20.4	61.4	245486	12	AC125575	AC125575 Rattus no	C 212	20.2	60.2	170984	5	AL445312	Continuation (9 of
C 140	20.4	61.4	246499	12	AC140754	AC140754 Rattus no	C 213	20.2	60.2	171272	5	AC146225	Continuation (9 of
C 141	20.4	61.4	251382	12	AC162260	AC162260 Bos tauru	C 214	20.2	60.2	173239	5	AC016399	Continuation (9 of
C 142	20.4	61.4	263879	12	AC096060	AC096060 Rattus no	C 215	20.2	60.2	175086	12	AC139609	Continuation (9 of
C 143	20.4	61.4	266671	12	AC096126	AC096126 Rattus no	C 216	20.2	60.2	177206	11	CR354558	Continuation (9 of
C 144	20.4	61.4	293904	12	AC165519	AC165519 Bos tauru	C 217	20.2	60.2	179603	2	CQ870485	Continuation (9 of
C 145	20.4	61.4	349887	15	BX527749	BX527749 Corynebact	C 218	20.2	60.2	183540	12	AC026114	Continuation (9 of
C 146	20.4	61.4	349980	2	AX127144	AX127144 Sequence	C 219	20.2	60.2	185530	12	AC096875	Continuation (9 of
C 147	20.2	60.8	428	11	AY844405	AY844405 Hyla cavi	C 220	20.2	60.2	188768	11	BX908760	Continuation (9 of
C 148	20.2	60.8	52539	5	AL627447	AL627447 Human DNA	C 221	20.2	60.2	188972	5	AC020611	Continuation (9 of
C 149	20.2	60.8	110000	12	AL359456_3	Continuation (4 of	C 222	20.2	60.2	189791	12	AC152354	Continuation (9 of
C 150	20.2	60.8	110000	15	CR628336_22	Continuation (23 o	C 223	20.2	60.2	190762	12	AC163191	Continuation (9 of
C 151	20.2	60.8	110000	15	CR628337_22	Continuation (23 o	C 224	20.2	60.2	193375	12	AC172341	Continuation (9 of
C 152	20.2	60.8	110000	15	AE017354_22	Continuation (23 o	C 225	20.2	60.2	193551	11	BX649283	Continuation (9 of
C 153	20.2	60.8	129835	12	AC177867	AC177867 Zea mays	C 226	20.2	60.2	198254	12	AC147089	Continuation (9 of
C 154	20.2	60.8	130808	11	CR392029	CR392029 Zebrafish	C 227	20.2	60.2	199367	12	AC160727	Continuation (9 of
C 155	20.2	60.8	142116	12	AC157612	AC157612 Ornithorh	C 228	20.2	60.2	199668	12	AC163957	Continuation (9 of
C 156	20.2	60.8	143579	12	AC157845	AC157845 Ornithorh	C 229	20.2	60.2	204328	12	AL591119	Continuation (9 of
C 157	20.2	60.8	145729	5	AC117489	AC117489 Homo sapi	C 230	20.2	60.2	208067	5	AC146508	Continuation (9 of
C 158	20.2	60.8	168729	12	AC087418	AC087418 Homo sapi	C 231	20.2	60.2	208764	12	AC095520	Continuation (9 of
C 159	20.2	60.8	172412	12	AC134351	AC134351 Rattus no	C 232	20.2	60.2	209228	11	BX890613	Continuation (9 of
C 160	20.2	60.8	172770	12	AC148425	AC148425 Ornithorh	C 233	20.2	60.2	209921	12	AC163135	Continuation (9 of
C 161	20.2	60.8	182439	12	AL607037	AL607037 Homo sapi	C 234	20.2	60.2	212371	12	CR936381	Continuation (9 of
C 162	20.2	60.8	182639	5	AC104307	AC104307 Homo sapi	C 235	20.2	60.2	212429	11	BX571828	Continuation (9 of
C 163	20.2	60.8	187159	5	AC090941	AC090941 Homo sapi	C 236	20.2	60.2	236484	11	BX539341	Continuation (9 of
C 164	20.2	60.8	205754	12	AC097001	AC097001 Rattus no	C 237	20.2	60.2	240052	12	AC169766	Continuation (9 of

238	20	60.2	30066	15	AE017159	AE017159 Chlamydom	311	19.8	59.6	245698	12	AC124841	AC124841 Rattus no
239	20	60.2	317233	12	AC130067	AC130067 Rattus no	312	19.8	59.6	246029	12	AC168577	AC168577 Strongylo
240	20	60.2	349980	2	C0870484	C0870484 Sequence	c 313	19.8	59.6	246613	12	AC126851	AC126851 Rattus no
241	19.8	59.6	289	7	G89085	G89085 S208F6370FC	c 314	19.8	59.6	253631	12	AC126155	AC126155 Rattus no
242	19.8	59.6	428	11	AY844377	AY844377 Duellmano	c 315	19.8	59.6	256998	12	AC115751	AC115751 Mus muscu
243	19.8	59.6	428	11	AY844388	AY844388 Hyla anne	c 316	19.8	59.6	259730	6	AC094880	AC094880 Rattus no
244	19.8	59.6	428	11	AY844389	AY844389 Hyla arbo	317	19.8	59.6	262632	12	AC094326	AC094326 Rattus no
245	19.8	59.6	428	11	AY844391	AY844391 Hyla aren	318	19.8	59.6	275706	12	AC120591	AC120591 Rattus no
246	19.8	59.6	428	11	AY844420	AY844420 Hyla japo	c 319	19.8	59.6	293854	12	AC126646	AC126646 Rattus no
247	19.8	59.6	428	11	AY844462	AY844462 Hyla equi	c 320	19.8	59.6	299420	12	AC094449	AC094449 Rattus no
248	19.8	59.6	428	11	AY844465	AY844465 Hyla vers	321	19.8	59.6	299876	12	AC106196	AC106196 Rattus no
249	19.8	59.6	428	11	AY844466	AY844466 Hyla walk	c 322	19.8	59.6	340000	5	AP001684	AP001684 Homo sapi
250	19.8	59.6	428	11	AY844518	AY844518 Scinax el	c 323	19.6	59.0	356	2	CS066201	CS066201 Sequence
251	19.8	59.6	428	11	AY844519	AY844519 Scinax fu	324	19.6	59.0	428	11	AY844358	AY844358 Acris cre
252	19.8	59.6	1509	11	AY583347	AY583347 Chigloss	325	19.6	59.0	428	11	AY844359	AY844359 Acris gry
253	19.8	59.6	1512	11	AY571662	AY571662 Hyla meri	326	19.6	59.0	428	11	AY844434	AY844434 Hyla mixe
254	19.8	59.6	1855	4	AK072046	AK072046 Oryza sat	327	19.6	59.0	1005	2	AR547935	AR547935 Sequence
255	19.8	59.6	1855	4	AK072046	AK072046 Oryza sat	c 328	19.6	59.0	1423	4	BT013914	BT013914 Lycopersi
256	19.8	59.6	2002	4	AK105688	AK105688 Oryza sat	c 329	19.6	59.0	2300	4	AF253056	AF253056 Candida a
257	19.8	59.6	2023	4	AK065632	AK065632 Oryza sat	c 330	19.6	59.0	2454	15	AB078638	AB078638 Helicobac
258	19.8	59.6	2500	4	CCY11081	Y11081 C. cinereus	c 331	19.6	59.0	2454	15	AB078638	AB078638 Helicobac
259	19.8	59.6	16745	6	AY257203	AY257203 Mus muscu	332	19.6	59.0	3031	4	AB032082	AB032082 Candida a
260	19.8	59.6	44455	4	YSCP9513	Y51033 Saccharomyc	c 333	19.6	59.0	26359	5	AL512268	AL512268 Human DNA
261	19.8	59.6	51621	5	AP000657	AP000657 Homo sapi	c 334	19.6	59.0	62288	5	AC005571	AC005571 Homo sapi
262	19.8	59.6	64829	12	AC173369	AC173369 Strongylo	335	19.6	59.0	68229	5	AC007090	AC007090 Homo sapi
263	19.8	59.6	71927	12	AC113149	AC113149 Homo sapi	336	19.6	59.0	94001	13	AC171141	AC171141 Helobdell
264	19.8	59.6	73442	12	AC1330692	AC1330692 Homo sapi	337	19.6	59.0	98399	5	AL162291	AL162291 Human DNA
265	19.8	59.6	93409	12	CT573813	CT573813 Danio rer	c 338	19.6	59.0	100000	5	AB020865	AB020865 Homo sapi
266	19.8	59.6	104374	11	EX649313	EX649313 Zebrafish	339	19.6	59.0	101377	12	AC152053	AC152053 Medicago
267	19.8	59.6	110000	4	AP008221	Continuation (111	340	19.6	59.0	109739	12	AC115437	Continuation (4 of
268	19.8	59.6	110000	13	AC165428	Continuation (3 of	341	19.6	59.0	110000	4	AP008221	Continuation (197
269	19.8	59.6	113812	4	AC149135	AC149135 Medicago	342	19.6	59.0	110000	15	CR522870	Continuation (34 o
270	19.8	59.6	123291	5	AC104070	AC104070 Homo sapi	343	19.6	59.0	110000	15	BA000011	Continuation (3 of
271	19.8	59.6	129206	12	HS03230H6	AL121940 Homo sapi	344	19.6	59.0	111446	4	AC143341	AC143341 Medicago
272	19.8	59.6	142834	4	OSJN00253	AL731611 Oryza sat	345	19.6	59.0	117329	4	AC148177	AC148177 Triticum
273	19.8	59.6	143820	12	AC135843	AC135843 Felis cat	c 346	19.6	59.0	121991	12	AM050677	AM050677 Triticum
274	19.8	59.6	146559	5	AC013262	AC013262 Homo sapi	c 347	19.6	59.0	129756	5	HS875H10	AL034350 Human DNA
275	19.8	59.6	155476	6	AC154816	AC154816 Mus muscu	c 348	19.6	59.0	136586	14	AC157604	AC157604 Ornithorh
276	19.8	59.6	157426	5	AC018449	AC018449 Homo sapi	349	19.6	59.0	143722	4	AP005723	AP005723 Oryza sat
277	19.8	59.6	159393	5	AC160568	AC160568 Pan trogl	350	19.6	59.0	144712	12	AC133559	AC133559 Rattus no
278	19.8	59.6	161475	5	HS164C20	AL090929 Human DNA	351	19.6	59.0	150549	4	AP004794	AP004794 Oryza sat
279	19.8	59.6	161566	12	AC136462	AC136462 Rattus no	352	19.6	59.0	153381	12	AP001990	AP001990 Homo sapi
280	19.8	59.6	162977	12	AC177178	AC177178 Strongylo	353	19.6	59.0	156929	12	AC026854	AC026854 Homo sapi
281	19.8	59.6	166285	12	AC174839	AC174839 Echinosps	c 354	19.6	59.0	157336	12	AC079787	AC079787 Homo sapi
282	19.8	59.6	170364	12	AC178856	AC178856 Strongylo	c 355	19.6	59.0	158794	12	AC023961	AC023961 Homo sapi
283	19.8	59.6	171656	12	AC178941	AC178941 Strongylo	c 356	19.6	59.0	159403	5	AC146155	AC146155 Pan trogl
284	19.8	59.6	174540	6	AC119900	AC119900 Mus muscu	357	19.6	59.0	160223	5	AC027117	AC027117 Homo sapi
285	19.8	59.6	174772	12	AC168491	AC168491 Strongylo	358	19.6	59.0	160835	5	AC087273	AC087273 Homo sapi
286	19.8	59.6	175339	5	AC005772	AC005772 Homo sapi	359	19.6	59.0	161194	12	AC018945	AC018945 Homo sapi
287	19.8	59.6	175840	12	AC170828	AC170828 Bos tauru	360	19.6	59.0	161675	12	AP001975	AP001975 Homo sapi
288	19.8	59.6	175859	12	AC161486	AC161486 Mus muscu	c 361	19.6	59.0	161754	5	AC093206	AC093206 Homo sapi
289	19.8	59.6	180693	12	AC178594	AC178594 Strongylo	c 362	19.6	59.0	162515	5	AC107032	AC107032 Homo sapi
290	19.8	59.6	180851	12	AC090210	AC090210 Homo sapi	363	19.6	59.0	168468	15	AC005510	AC005510 Homo sapi
291	19.8	59.6	181211	12	AP001904	AP001904 Homo sapi	364	19.6	59.0	173118	12	AC109295	AC109295 Mus muscu
292	19.8	59.6	183385	6	AC117206	AC117206 Mus muscu	365	19.6	59.0	175523	6	CT025565	CT025565 Mouse DNA
293	19.8	59.6	188083	6	AC127309	AC127309 Mus muscu	c 366	19.6	59.0	176036	12	AM050683	AM050683 Triticum
294	19.8	59.6	188616	12	AC156152	AC156152 Bos tauru	c 367	19.6	59.0	176804	12	AC053484	AC053484 Homo sapi
295	19.8	59.6	200728	12	AC131357	AC131357 Rattus no	c 368	19.6	59.0	177244	12	AC123288	AC123288 Rattus no
296	19.8	59.6	201237	12	AC111505	AC111505 Rattus no	369	19.6	59.0	179680	12	AL669953	AL669953 Mouse DNA
297	19.8	59.6	203441	12	AC167131	AC167131 Mus muscu	370	19.6	59.0	179956	12	AC094865	AC094865 Rattus no
298	19.8	59.6	206246	12	AC132970	AC132970 Rattus no	371	19.6	59.0	180924	6	AC164175	AC164175 Mus muscu
299	19.8	59.6	207088	6	AC160635	AC160635 Mus muscu	c 372	19.6	59.0	181220	6	AC136008	AC136008 Mus muscu
300	19.8	59.6	211343	6	AC122630	AC122630 Rattus no	c 373	19.6	59.0	185724	12	AC146285	AC146285 Callicebu
301	19.8	59.6	213552	5	AC010928	AC010928 Homo sapi	c 374	19.6	59.0	186744	12	AC176383	AC176383 Strongylo
302	19.8	59.6	215020	12	AC170695	AC170695 Bos tauru	c 375	19.6	59.0	186901	12	AF267167	AF267167 Homo sapi
303	19.8	59.6	222032	12	AC096942	AC096942 Rattus no	c 376	19.6	59.0	190735	12	AC026084	AC026084 Homo sapi
304	19.8	59.6	222745	12	AC127100	AC127100 Rattus no	c 377	19.6	59.0	192992	12	AC145957	AC145957 Gallus ga
305	19.8	59.6	224334	12	AC096041	AC096041 Rattus no	c 378	19.6	59.0	193614	6	AC110219	AC110219 Mus muscu
306	19.8	59.6	231639	12	AC179487	AC179487 Strongylo	379	19.6	59.0	193986	14	AC124157	AC124157 Equus cab
307	19.8	59.6	234462	12	AC109738	AC109738 Rattus no	c 380	19.6	59.0	195290	5	AP006249	AP006249 Homo sapi
308	19.8	59.6	235452	12	AC128416	AC128416 Rattus no	c 381	19.6	59.0	195412	6	AL772387	AL772387 Mouse DNA
309	19.8	59.6	235468	6	AC105586	AC105586 Rattus no	c 382	19.6	59.0	196132	6	AC165082	AC165082 Mus muscu
310	19.8	59.6	237453	12	CR392369	CR392369 Danio rer	c 383	19.6	59.0	209937	6	AC138192	AC138192 Mus muscu

C 384	19.6	59.0	209945	12	AC139756	AC139756 Mus muscu	C 457	19.4	58.4	164040	12	AC147530	Otolemur
C 385	19.6	59.0	211300	12	CR790363	CR790363 Dantio rer	C 458	19.4	58.4	164175	12	AC119056	Papio anu
C 386	19.6	59.0	214665	6	AL646046	AL646046 Mouse DNA	C 459	19.4	58.4	165496	5	AC119056	Homo sapi
C 387	19.6	59.0	221115	12	AC093242	AC093242 Homo sapi	C 460	19.4	58.4	166118	14	CT009560	Pig DNA s
C 388	19.6	59.0	221708	12	AC103447	AC103447 Rattus no	C 461	19.4	58.4	170979	12	AC118896	Rattus no
C 389	19.6	59.0	221953	12	AC130078	AC130078 Rattus no	C 462	19.4	58.4	171585	12	AC148301	Pan trogl
C 390	19.6	59.0	224373	12	AC129363	AC129363 Rattus no	C 463	19.4	58.4	171806	6	AC113510	Mus muscu
C 391	19.6	59.0	226029	12	AC171007	AC171007 Gallus ga	C 464	19.4	58.4	171822	12	CT025616	Mus muscu
C 392	19.6	59.0	230703	12	AC120663	AC120663 Rattus no	C 465	19.4	58.4	172835	12	AC121174	Rattus no
C 393	19.6	59.0	233112	12	AC172238	AC172238 Bos tauru	C 466	19.4	58.4	174545	12	AC176158	Strongylo
C 394	19.6	59.0	242151	12	AC159523	AC159523 Bos tauru	C 467	19.4	58.4	174582	12	AC118488	Rattus no
C 395	19.6	59.0	247511	12	AC096432	AC096432 Rattus no	C 468	19.4	58.4	175485	5	AC147539	Pan trogl
C 396	19.6	59.0	257513	12	AC127093	AC127093 Rattus no	C 469	19.4	58.4	175504	12	AC153161	Glycine m
C 397	19.6	59.0	270120	12	AC128627	AC128627 Rattus no	C 470	19.4	58.4	176241	12	AC167936	Sus scrof
C 398	19.6	59.0	271200	12	AC131538	AC131538 Rattus no	C 471	19.4	58.4	176861	5	AC097065	Homo sapi
C 399	19.6	59.0	271233	12	AC096597	AC096597 Rattus no	C 472	19.4	58.4	178044	12	AC121663	Rattus no
C 400	19.6	59.0	275292	12	AC129270	AC129270 Rattus no	C 473	19.4	58.4	178246	12	AC124162	Rattus no
C 401	19.6	59.0	275443	12	AC096407	AC096407 Rattus no	C 474	19.4	58.4	178872	12	CT573008	Dantio rer
C 402	19.6	59.0	276163	12	AC130878	AC130878 Rattus no	C 475	19.4	58.4	180072	5	AC017092	Homo sapi
C 403	19.6	59.0	283341	12	AC128766	AC128766 Rattus no	C 476	19.4	58.4	180150	5	AC118138	Homo sapi
C 404	19.6	59.0	294915	12	AC155005	AC155005 Bos tauru	C 477	19.4	58.4	180161	12	AC180837	Strongylo
C 405	19.6	59.0	295001	12	AC118175	AC118175 Rattus no	C 478	19.4	58.4	181234	12	AC167935	Sus scrof
C 406	19.6	59.0	309922	12	AC119578	AC119578 Rattus no	C 479	19.4	58.4	182902	5	AC092405	Papio anu
C 407	19.6	59.0	318862	12	AC109547	AC109547 Rattus no	C 480	19.4	58.4	183118	12	AC160722	Actus nan
C 408	19.4	58.4	312	2	CQ657729	CQ657729 Sequence	C 481	19.4	58.4	183412	12	AP001899	Homo sapi
C 409	19.4	58.4	428	11	AY844487	AY844487 Osteopllu	C 482	19.4	58.4	184164	5	AC091020	Homo sapi
C 410	19.4	58.4	559	7	G09062	G09062 human STS C	C 483	19.4	58.4	184951	6	AC155292	Mus muscu
C 411	19.4	58.4	660	1	AY214920	AY214920 Uncultu	C 484	19.4	58.4	184951	5	AF152363	Homo sapi
C 412	19.4	58.4	987	11	AY543804	AY543804 Salmo sal	C 485	19.4	58.4	185763	5	AF152363	Homo sapi
C 413	19.4	58.4	992	2	CQ872398	CQ872398 Sequence	C 486	19.4	58.4	189108	12	AC178034	Strongylo
C 414	19.4	58.4	1056	7	BV176987	BV176987 sqm91960	C 487	19.4	58.4	189281	12	AC012458	Homo sapi
C 415	19.4	58.4	1957	11	AY789437S3	AY789439 Oncorhync	C 488	19.4	58.4	190363	5	AC000052	Homo sapi
C 416	19.4	58.4	2259	5	AF410455	AF410455 Homo sapi	C 489	19.4	58.4	191041	12	AC024114	Mus muscu
C 417	19.4	58.4	2321	2	AX086673	AX086673 Sequence	C 490	19.4	58.4	191434	5	AC087477	Homo sapi
C 418	19.4	58.4	2322	5	BC019008	BC019008 Homo sapi	C 491	19.4	58.4	192653	5	AC124781	Homo sapi
C 419	19.4	58.4	2324	5	BC007823	BC007823 Homo sapi	C 492	19.4	58.4	193015	5	AC006460	Homo sapi
C 420	19.4	58.4	2326	2	AR578216	AR578216 Sequence	C 493	19.4	58.4	193456	12	AC149228	Pan trogl
C 421	19.4	58.4	2737	4	AK111149	AK111149 Oryza sat	C 494	19.4	58.4	197483	5	AC104164	Homo sapi
C 422	19.4	58.4	3738	2	AR376414	AR376414 Sequence	C 495	19.4	58.4	198488	5	AC023794	Homo sapi
C 423	19.4	58.4	8271	5	AF489699	AF489699 Homo sapi	C 496	19.4	58.4	200061	6	AC153535	Mus muscu
C 424	19.4	58.4	11936	15	AE005946	AE005946 Caulobact	C 497	19.4	58.4	200355	12	AC164743	Bos tauru
C 425	19.4	58.4	15229	12	AM182498	AM182498 Equus cab	C 498	19.4	58.4	201376	12	AC164743	Bos tauru
C 426	19.4	58.4	46077	12	AC137289	AC137289 Rattus no	C 499	19.4	58.4	201397	5	BS000130	Pan trogl
C 427	19.4	58.4	58143	12	AC137177	AC137177 Rattus no	C 500	19.4	58.4	202412	6	AL670673	Mouse DNA
C 428	19.4	58.4	79943	5	AL353664	AL353664 Human DNA	C 501	19.4	58.4	205118	12	AC118478	Mus muscu
C 429	19.4	58.4	80589	11	BX470257	BX470257 Zebrafish	C 502	19.4	58.4	205174	5	AC011804	Homo sapi
C 430	19.4	58.4	81896	5	AP001614	AP001614 Homo sapi	C 503	19.4	58.4	207961	12	AC111825	Rattus no
C 431	19.4	58.4	90526	5	AC092393	AC092393 Homo sapi	C 504	19.4	58.4	211056	6	AL591207	Mouse DNA
C 432	19.4	58.4	104980	12	CT030899	CT030699 Dantio rer	C 505	19.4	58.4	212548	6	AC102372	Mus muscu
C 433	19.4	58.4	110000	4	CR382139_08	Continuation (9 of	C 506	19.4	58.4	213033	12	AC131482	Rattus no
C 434	19.4	58.4	110000	4	AP008209_268	Continuation (269	C 507	19.4	58.4	219053	12	AC167784	Glycine m
C 435	19.4	58.4	110000	12	AC109413_0	AC109413 Rattus no	C 508	19.4	58.4	219256	5	AP001743	Homo sapi
C 436	19.4	58.4	110000	12	AC112029_0	AC112029 Rattus no	C 509	19.4	58.4	222161	12	AC147935	Otolemur
C 437	19.4	58.4	110000	12	AC112029_1	Continuation (2 of	C 510	19.4	58.4	224058	5	AC008513	Homo sapi
C 438	19.4	58.4	110000	15	CP000077_08	Continuation (9 of	C 511	19.4	58.4	225829	6	AC129609	Mus muscu
C 439	19.4	58.4	118327	4	AC045521	AC004521 Arabidops	C 512	19.4	58.4	228117	12	AC170832	Bos tauru
C 440	19.4	58.4	118362	12	AC168753	AC168753 Strongylo	C 513	19.4	58.4	234607	12	AC152095	Bos tauru
C 441	19.4	58.4	121661	13	AC150385	AC150385 Branchios	C 514	19.4	58.4	236720	12	AC123083	Rattus no
C 442	19.4	58.4	124706	11	BX927312	BX927312 Zebrafish	C 515	19.4	58.4	238440	12	AC137383	Rattus no
C 443	19.4	58.4	135931	4	AC079736	AC079736 Oryza sat	C 516	19.4	58.4	239242	12	AC099428	Rattus no
C 444	19.4	58.4	141710	12	AC170079	AC170079 Ixodonta	C 517	19.4	58.4	240673	12	AC181744	Strongylo
C 445	19.4	58.4	143527	6	AC115044	AC115044 Mus muscu	C 518	19.4	58.4	249433	12	AC051620	Mus muscu
C 446	19.4	58.4	151207	6	AL604025	AL604025 Mouse DNA	C 519	19.4	58.4	260409	5	AC004019	Homo sapi
C 447	19.4	58.4	152033	5	AC027319	AC027319 Homo sapi	C 520	19.4	58.4	261376	12	AC098996	Rattus no
C 448	19.4	58.4	152413	12	AC181948	AC181948 Strongylo	C 521	19.4	58.4	265169	12	AC110476	Rattus no
C 449	19.4	58.4	156207	12	AC025507	AC025507 Homo sapi	C 522	19.4	58.4	272400	12	AC110146	Rattus no
C 450	19.4	58.4	157430	6	AC127678	AC127678 Mus muscu	C 523	19.4	58.4	276015	12	CR759962	Homo sapi
C 451	19.4	58.4	161594	6	AC157928	AC157928 Mus muscu	C 524	19.4	58.4	276913	12	CR759962	Homo sapi
C 452	19.4	58.4	161005	11	BX537351	BX537351 Zebrafish	C 525	19.4	58.4	283299	12	AC112576	Rattus no
C 453	19.4	58.4	161406	12	AC121326	AC121326 Homo sapi	C 526	19.4	58.4	287233	12	AC110639	Rattus no
C 454	19.4	58.4	163009	5	AC011894	AC011894 Homo sapi	C 527	19.4	58.4	306522	12	AC137295	Rattus no
C 455	19.4	58.4	163503	11	BX927205	BX927205 Zebrafish	C 528	19.4	58.4	349635	12	AC093978	Rattus no
C 456	19.4	58.4	163612	12	AC011794	AC011794 Homo sapi	C 529	19.2	57.8	428	11	AY844413	Hyla cycl

530	19.2	57.8	428	11	AY844438	AY844438 Hyla neph	c 603	19.2	57.8	31268	13	CEH08M01	293384 Caenorhabdi
c 531	19.2	57.8	439	7	BV292436	BV292436 S232P6431	604	19.2	57.8	31434	13	CEY68A4A	AL021503 Caenorhab
c 532	19.2	57.8	620	7	BV048212	BV048212 S212P6044	605	19.2	57.8	32009	13	CET06D8	249130 Caenorhabdi
c 533	19.2	57.8	664	11	AY609297	AY609297 Chlorospi	606	19.2	57.8	32488	13	AF100307	AF100307 Caenorhab
c 534	19.2	57.8	668	7	BV388502	BV388502 S244P617R	607	19.2	57.8	32591	13	CEFI6D3	278062 Caenorhabdi
c 535	19.2	57.8	673	11	AY609296	AY609296 Chlorospi	608	19.2	57.8	32838	13	AF003385	AF003385 Caenorhab
c 536	19.2	57.8	689	7	BV655500	BV655500 S216P6015	609	19.2	57.8	32960	13	CET27E5	282284 Caenorhabdi
c 537	19.2	57.8	829	11	AF109418	AF109418 Icterus m	610	19.2	57.8	33802	13	CER07H7	281107 Caenorhabdi
c 538	19.2	57.8	842	11	AF382968	AF382968 Seiurus a	611	19.2	57.8	34073	13	CEFI1D11	281500 Caenorhabdi
c 539	19.2	57.8	998	11	AF447354	AF447354 Seiurus a	612	19.2	57.8	34494	13	AC006633	AC006633 Caenorhab
c 540	19.2	57.8	1001	2	AK582469	AK582469 Sequence	613	19.2	57.8	34494	13	AC006633	AC006633 Caenorhab
c 541	19.2	57.8	1293	4	AK106009	AK106009 Oryza sat	614	19.2	57.8	35061	13	U47144	U47144 Caenorhabdi
c 542	19.2	57.8	1599	6	BC034063	BC034063 Mus muscu	615	19.2	57.8	35077	13	CEY38H8A	AL021483 Caenorhab
c 543	19.2	57.8	1670	11	AF468612	AF468612 Alimophila	616	19.2	57.8	35950	13	U49947	U49947 Caenorhabdi
c 544	19.2	57.8	1671	11	AF468611	AF468611 Alimophila	617	19.2	57.8	35950	13	CER07H5	250874 Caenorhabdi
c 545	19.2	57.8	1704	13	AK115593	AK115593 Clona inc	618	19.2	57.8	36170	13	CER10E4	250874 Caenorhabdi
c 546	19.2	57.8	1716	10	DQ228931	DQ228931 Newcastle	619	19.2	57.8	36344	13	U58728	U58728 Caenorhabdi
c 547	19.2	57.8	1716	10	DQ234586	DQ234586 Newcastle	620	19.2	57.8	36429	13	U29380	U29380 Caenorhabdi
c 548	19.2	57.8	1716	10	DQ234587	DQ234587 Newcastle	621	19.2	57.8	37427	13	CEH12119	AF039049 Caenorhab
c 549	19.2	57.8	1801	10	DQ023556	DQ023556 Newcastle	622	19.2	57.8	38375	13	CEP56D5	298851 Caenorhabdi
c 550	19.2	57.8	1828	2	AK698569	AK698569 Sequence	623	19.2	57.8	38861	13	CEB06B8	269662 Caenorhabdi
c 551	19.2	57.8	1828	2	AK338236	AK338236 Sequence	624	19.2	57.8	39333	13	CEB0564	281463 Caenorhabdi
c 552	19.2	57.8	1899	13	AF478698	AF478698 Drosophil	625	19.2	57.8	39902	13	CEB03H4	281492 Caenorhabdi
c 553	19.2	57.8	1899	13	AF478699	AF478699 Drosophil	626	19.2	57.8	40094	13	CEP44F4	237092 Caenorhabdi
c 554	19.2	57.8	1899	13	AF478700	AF478700 Drosophil	627	19.2	57.8	40145	13	CEC36A4	266495 Caenorhabdi
c 555	19.2	57.8	1936	6	BC036174	BC036174 Mus muscu	628	19.2	57.8	40641	13	AC006816	AC006816 Caenorhab
c 556	19.2	57.8	1971	13	AY084098	AY084098 Drosophil	629	19.2	57.8	41299	13	CEC06H5	292775 Caenorhabdi
c 557	19.2	57.8	2090	2	CQ588981	CQ588981 Sequence	630	19.2	57.8	41811	13	U58746	U58746 Caenorhabdi
c 558	19.2	57.8	2102	6	BC039217	BC039217 Mus muscu	631	19.2	57.8	42297	13	CEC33A12	268493 Caenorhabdi
c 559	19.2	57.8	2142	13	AF478697	AF478697 Drosophil	632	19.2	57.8	42300	13	U80452	U80452 Caenorhabdi
c 560	19.2	57.8	2166	13	AY071591	AY071591 Drosophil	633	19.2	57.8	42574	13	AF039053	AF039053 Caenorhab
c 561	19.2	57.8	2174	6	BC036173	BC036173 Mus muscu	634	19.2	57.8	42889	13	U10438	U10438 Caenorhabdi
c 562	19.2	57.8	2226	13	AF158747	AF158747 Drosophil	635	19.2	57.8	42968	13	CEY37A1A	299270 Caenorhabdi
c 563	19.2	57.8	2321	6	RATBEM3	D45414 Rat mRNA fo	636	19.2	57.8	43525	13	U88311	U88311 Caenorhabdi
c 564	19.2	57.8	2420	6	AY079517S2	AY079517 Mus muscu	637	19.2	57.8	44352	13	AF098501	AF098501 Caenorhab
c 565	19.2	57.8	2496	2	AK447985	AK447985 Sequence	638	19.2	57.8	44437	13	AC006645	AC006645 Caenorhab
c 566	19.2	57.8	2952	6	RNICA105	X92563 R. norvegici	639	19.2	57.8	44614	13	CEC29B6	272504 Caenorhabdi
c 567	19.2	57.8	3358	4	AK065594	AK065594 Oryza sat	640	19.2	57.8	51235	13	AC006769	AC006769 Caenorhab
c 568	19.2	57.8	3397	6	RNU40652	U40652 Rattus norv	641	19.2	57.8	55613	5	AL512443	AL512443 Human DNA
c 569	19.2	57.8	3477	6	RATPDPFLP	D38222 Rattus sp.	642	19.2	57.8	65181	12	AC087443	AC087443 Homo sapi
c 570	19.2	57.8	4090	2	CQ588980	CQ588980 Sequence	643	19.2	57.8	65181	12	AC087443	AC087443 Homo sapi
c 571	19.2	57.8	4691	11	AY650211	AY650211 Seiurus a	644	19.2	57.8	67943	13	AC006722	AC006722 Caenorhab
c 572	19.2	57.8	4694	11	AY650193	AY650193 Dendroica	645	19.2	57.8	68270	13	CEY7A5A	AL034489 Caenorhab
c 573	19.2	57.8	8020	13	AC024832	AC024832 Caenorhab	646	19.2	57.8	69490	5	AC114294	AC114294 Homo sapi
c 574	19.2	57.8	8625	15	AF486545	AF486545 Campyloba	647	19.2	57.8	76022	12	AC168920	AC168920 Bos tauru
c 575	19.2	57.8	10048	13	CET26H8	282057 Caenorhabdi	648	19.2	57.8	83823	12	AC006789	AC006789 Caenorhab
c 576	19.2	57.8	10186	15	AF486550	AF486550 Campyloba	649	19.2	57.8	83943	2	CQ870148	CQ870148 Sequence
c 577	19.2	57.8	10187	15	AF486544	AF486544 Campyloba	650	19.2	57.8	86290	5	AC112503	AC112503 Homo sapi
c 578	19.2	57.8	10187	15	AF486551	AF486551 Campyloba	651	19.2	57.8	87428	13	CEY64G10A	AL110498 Caenorhab
c 579	19.2	57.8	10187	15	AF486558	AF486558 Campyloba	652	19.2	57.8	91927	5	AC004771	AC004771 Homo sapi
c 580	19.2	57.8	10627	2	I13706	I13706 Sequence 12	653	19.2	57.8	92817	5	AL591202	AL591202 Human DNA
c 581	19.2	57.8	12222	2	CS098060	CS098060 Sequence	654	19.2	57.8	94486	11	AC149069	AC149069 X. tropic
c 582	19.2	57.8	12222	2	AK390794	AK390794 Sequence	655	19.2	57.8	94762	12	AF157691	AF157691 Xenopus t
c 583	19.2	57.8	12222	5	HUMAIATP	R02212 Human alpha	656	19.2	57.8	95069	4	AP006265	AP006265 Oryza sat
c 584	19.2	57.8	14661	5	AY795074	AY795074 Homo sapi	657	19.2	57.8	97675	12	AC148534	AC148534 Macaca mu
c 585	19.2	57.8	15064	13	AC006713	AC006713 Caenorhab	658	19.2	57.8	98348	12	AC158848	AC158848 Bos tauru
c 586	19.2	57.8	15093	13	AF003149	AF003149 Caenorhab	659	19.2	57.8	99083	13	CEY37A1B	AL023835 Caenorhab
c 587	19.2	57.8	15106	6	AF521697	AF521697 Mus muscu	660	19.2	57.8	100883	12	AC157686	AC157686 Xenopus t
c 588	19.2	57.8	18572	13	CEFA7G9	274035 Caenorhabdi	661	19.2	57.8	104970	13	AC024877	AC024877 Caenorhab
c 589	19.2	57.8	18806	13	CEY45F10C	299273 Caenorhabdi	662	19.2	57.8	110000	4	AP008213_070	Continuation (71 o
c 590	19.2	57.8	22528	13	CEZK1225	AL022289 Caenorhab	663	19.2	57.8	110000	4	AP008214_274	Continuation (275
c 591	19.2	57.8	24171	13	CEC31A11	283218 Caenorhabdi	664	19.2	57.8	110000	4	AP008217_210	Continuation (211
c 592	19.2	57.8	24330	13	AC024800	AC024800 Caenorhab	665	19.2	57.8	110000	4	AP008217_219	Continuation (220
c 593	19.2	57.8	24919	13	AF098991	AF098991 Caenorhab	666	19.2	57.8	110000	4	AP008218_085	Continuation (66 o
c 594	19.2	57.8	24950	13	AF014939	AF014939 Caenorhab	667	19.2	57.8	110000	4	CR382121_00	CR382121 Kluyverom
c 595	19.2	57.8	25406	13	U42437	U42437 Caenorhabdi	668	19.2	57.8	110000	4	AE016815_7	Continuation (8 of
c 596	19.2	57.8	25595	13	AF039045	AF039045 Caenorhab	669	19.2	57.8	110000	4	AP008207_252	Continuation (253
c 597	19.2	57.8	25856	13	CEC14A6	281470 Caenorhabdi	670	19.2	57.8	110000	4	AP008207_370	Continuation (371
c 598	19.2	57.8	29096	13	CEC05C12	268333 Caenorhabdi	671	19.2	57.8	110000	4	AP008211_057	Continuation (58 o
c 599	19.2	57.8	29344	13	CEY116A8B	AL021469 Caenorhab	672	19.2	57.8	110000	12	AC150790_0	AC150790 Bos tauru
c 600	19.2	57.8	29655	13	U40061	U40061 Caenorhabdi	673	19.2	57.8	110000	12	AC150872_0	AC150872 Bos tauru
c 601	19.2	57.8	29977	13	AF025461	AF025461 Caenorhab	674	19.2	57.8	110000	12	CEY111B2_1	Continuation (2 of
c 602	19.2	57.8	30767	13	AC006631	AC006631 Caenorhab	675	19.2	57.8	110000	15	CP000082_12	Continuation (13 o

676	19.2	57.8	110000	15	CP000082_13	Continuation (14 o	749	19.2	57.8	178434	12	AC118102
677	19.2	57.8	110000	15	CR931997_09	Continuation (10 o	750	19.2	57.8	180069	12	AC017041
678	19.2	57.8	110000	15	CR954246_25	Continuation (26 o	751	19.2	57.8	180742	5	AL359753
679	19.2	57.8	110000	15	BA000028_05	Continuation (6 of	752	19.2	57.8	181513	12	AC147749
680	19.2	57.8	117629	12	DQ369747_05	DQ369747 Brassica	753	19.2	57.8	181914	12	AC024640
681	19.2	57.8	117919	4	AP003220	AP003220 Oryza sat	754	19.2	57.8	182356	6	AC154793
682	19.2	57.8	120984	5	AC006463	AC006463 Homo sapi	755	19.2	57.8	183046	2	CS086327
683	19.2	57.8	122855	13	AC084159	AC084159 Caenorhab	756	19.2	57.8	183099	12	AC034144
684	19.2	57.8	123496	12	AC178821	AC178821 Strongylo	757	19.2	57.8	183870	5	AC115284
685	19.2	57.8	123599	12	AC158852	AC158852 Bos tauru	758	19.2	57.8	184835	12	AC068091
686	19.2	57.8	124592	4	AC114828	AC114828 Oryza sat	759	19.2	57.8	186487	12	AC119833
687	19.2	57.8	129169	12	AC032038	AC032038 Homo sapi	760	19.2	57.8	186607	6	AL845292
688	19.2	57.8	131387	6	AC154403	AC154403 Mus muscu	761	19.2	57.8	187825	12	AC161061
689	19.2	57.8	134222	5	AB045363	AB045363 Homo sapi	762	19.2	57.8	187832	6	AC161596
690	19.2	57.8	137481	4	AP004704	AP004704 Oryza sat	763	19.2	57.8	188294	6	AC153608
691	19.2	57.8	138508	4	CNS08CCU	AL954156 Oryza sat	764	19.2	57.8	189610	5	AC118584
692	19.2	57.8	140620	4	AP003415	AP003415 Oryza sat	765	19.2	57.8	190628	12	AC178683
693	19.2	57.8	140788	12	AC148419	AC148419 Dasyapus n	766	19.2	57.8	190673	6	AL645808
694	19.2	57.8	141027	4	AC132486	AC132486 Oryza sat	767	19.2	57.8	193188	12	AC006884
695	19.2	57.8	142938	6	AC118735	AC118735 Mus muscu	768	19.2	57.8	193301	5	CNS01DRJ
696	19.2	57.8	145514	4	AP003681	AP003681 Oryza sat	769	19.2	57.8	193909	6	AC115797
697	19.2	57.8	145697	5	AC020642	AC020642 Homo sapi	770	19.2	57.8	196383	12	AC113159
698	19.2	57.8	146570	12	AC149792	AC149792 Aedes aeg	771	19.2	57.8	196947	6	AC121843
699	19.2	57.8	146927	12	AC020144	AC020144 Drosophil	772	19.2	57.8	198349	6	AC151298
700	19.2	57.8	147423	12	AC177333	AC177333 Strongylo	773	19.2	57.8	199299	6	AC131718
701	19.2	57.8	147505	5	CNS01D7A	AL132708 Human chr	774	19.2	57.8	199631	12	AC175342
702	19.2	57.8	148120	12	AC141805	AC141805 Apis mell	775	19.2	57.8	200958	4	AP002058
703	19.2	57.8	150789	12	AC006885	AC006885 Caenorhab	776	19.2	57.8	200968	4	AP005516
704	19.2	57.8	152272	4	AC137589	AC137589 Oryza sat	777	19.2	57.8	201031	6	AC154356
705	19.2	57.8	152272	4	AC144558	AC144558 Oryza sat	778	19.2	57.8	202719	12	AC111339
706	19.2	57.8	152637	12	AC006734	AC006734 Caenorhab	779	19.2	57.8	203950	12	AC127849
707	19.2	57.8	153388	4	AP003289	AP003289 Oryza sat	780	19.2	57.8	209480	12	AC140965
708	19.2	57.8	153587	12	AC148751	AC148751 Rhinolo	781	19.2	57.8	209583	6	AC165442
709	19.2	57.8	154482	6	AC126430	AC126430 Mus muscu	782	19.2	57.8	210196	5	AC016717
710	19.2	57.8	154482	6	AC136842	AC136842 Oryza sat	783	19.2	57.8	210558	5	AC016717
711	19.2	57.8	156312	5	AC138470	AC138470 Homo sapi	784	19.2	57.8	211910	6	AC116573
712	19.2	57.8	156321	2	CS162220	CS162220 Sequence	785	19.2	57.8	212071	6	AC136754
713	19.2	57.8	156331	5	AC074331	AC074331 Homo sapi	786	19.2	57.8	212309	6	AC154628
714	19.2	57.8	156342	5	AC158566	AC158566 Mus muscu	787	19.2	57.8	212682	12	AC157414
715	19.2	57.8	156958	5	AC013460	AC013460 Homo sapi	788	19.2	57.8	213401	12	AC124940
716	19.2	57.8	157722	13	AC092248	AC092248 Drosophil	789	19.2	57.8	213473	6	AC140287
717	19.2	57.8	159420	6	CT025589	CT025589 Mouse DNA	790	19.2	57.8	215664	12	AC129367
718	19.2	57.8	160163	12	AC180897	AC180897 Strongylo	791	19.2	57.8	216132	12	AC146533
719	19.2	57.8	160736	11	BX530031	BX530031 Zebrafish	792	19.2	57.8	216935	12	AC147486
720	19.2	57.8	161054	5	AC104444	AC104444 Homo sapi	793	19.2	57.8	218381	12	AC150564
721	19.2	57.8	162397	12	AC012536	AC012536 Homo sapi	794	19.2	57.8	222520	12	AC120336
722	19.2	57.8	162625	12	CET08B3	Z83124 Caenorhabd	795	19.2	57.8	222895	12	AC084065
723	19.2	57.8	163396	5	AL161445	AL161445 Human DNA	796	19.2	57.8	232962	12	AC094146
724	19.2	57.8	163528	6	AC121951	AC121951 Mus muscu	797	19.2	57.8	232310	6	AC123752
725	19.2	57.8	165871	5	CNS01DVO	AL135998 Human chr	798	19.2	57.8	224009	12	AC131058
726	19.2	57.8	167336	12	AC153669	AC153669 Bos tauru	799	19.2	57.8	224318	12	AC160377
727	19.2	57.8	167475	6	AC090881	AC090881 Mus Muscu	800	19.2	57.8	227606	13	CEY53F4B
728	19.2	57.8	167576	12	AC147297	AC147297 Pan trogl	801	19.2	57.8	227607	6	AC120570
729	19.2	57.8	167724	12	AC147577	AC147577 Orolemur	802	19.2	57.8	227940	12	AC151108
730	19.2	57.8	168821	6	AC138638	AC138638 Mus muscu	803	19.2	57.8	228943	12	AC152222
731	19.2	57.8	168946	11	CR847948	CR847948 Zebrafish	804	19.2	57.8	229867	12	AC091287
732	19.2	57.8	169226	12	AC006913	AC006913 Caenorhab	805	19.2	57.8	230058	12	AC108229
733	19.2	57.8	169246	5	AC091815	AC091815 Homo sapi	806	19.2	57.8	230879	5	AP003465
734	19.2	57.8	170094	5	AC146033	AC146033 Pan trogl	807	19.2	57.8	231778	11	BX511175
735	19.2	57.8	170164	12	CT573031	CT573031 Mus muscu	808	19.2	57.8	232415	12	AC099429
736	19.2	57.8	170993	12	AC021528	AC021528 Homo sapi	809	19.2	57.8	232571	6	AC116557
737	19.2	57.8	171141	6	AC157651	AC157651 Mus muscu	810	19.2	57.8	232830	12	CT033781
738	19.2	57.8	171787	5	AC100777	AC100777 Homo sapi	811	19.2	57.8	235045	12	AC119588
739	19.2	57.8	172942	12	AC016899	AC016899 Homo sapi	812	19.2	57.8	235071	12	AC115654
740	19.2	57.8	172959	12	AC119422	AC119422 Papio anu	813	19.2	57.8	235720	12	AC120712
741	19.2	57.8	173556	12	AC074388	AC074388 Homo sapi	814	19.2	57.8	235852	6	AC167244
742	19.2	57.8	173803	12	CNS01DV2	AL133448 Homo sapi	815	19.2	57.8	236144	13	AE003643
743	19.2	57.8	173992	12	CT030660	CT030660 Mus muscu	816	19.2	57.8	237374	6	AC112030
744	19.2	57.8	174298	12	AC167299	AC167299 Oryctolag	817	19.2	57.8	238722	12	AC098108
745	19.2	57.8	174642	5	AC096751	AC096751 Homo sapi	818	19.2	57.8	239591	12	AC128432
746	19.2	57.8	174790	12	AC181037	AC181037 Strongylo	819	19.2	57.8	240309	12	AC103118
747	19.2	57.8	175396	12	AC156621	AC156621 Mus muscu	820	19.2	57.8	240931	12	AC107097
748	19.2	57.8	176584	12	AC027812	AC027812 Homo sapi	821	19.2	57.8	245805	12	AC006752
749	19.2	57.8	176584	12	AC027812	AC027812 Homo sapi	821	19.2	57.8	245805	12	AC006752



C 822	19.2	57.8	248510	12	AC095814	AC095814 Rattus no	C 895	19	57.2	110000	15	AE017285_20	Continuation (21 o
C 823	19.2	57.8	249200	12	AC094799	AC094799 Rattus no	C 896	19	57.2	111973	12	AP007350	AP007350 Lotus cor
C 824	19.2	57.8	249262	12	AC079430	AC079430 Mus muscu	C 897	19	57.2	117217	5	AP002076	AP002076 Homo sapi
C 825	19.2	57.8	249481	12	AC159672	AC159672 Bos tauru	C 898	19	57.2	117752	5	AC004743	AC004743 Homo sapi
C 826	19.2	57.8	251742	12	AC096526	AC096526 Rattus no	C 899	19	57.2	125860	4	AP004151	AP004151 Oryza sat
C 827	19.2	57.8	252474	12	AC093999	AC093999 Rattus no	C 900	19	57.2	127006	5	AL138849	AL138849 Human DNA
C 828	19.2	57.8	254497	12	AC119520	AC119520 Rattus no	C 901	19	57.2	129690	6	AC138601	AC138601 Mus muscu
C 829	19.2	57.8	255305	12	AC160576	AC160576 Bos tauru	C 902	19	57.2	133878	12	AC136064	AC136064 Rattus no
C 830	19.2	57.8	256719	6	AC154873	AC154873 Mus muscu	C 903	19	57.2	140056	12	BS511131	BS511131 Danio rer
C 831	19.2	57.8	257158	12	AC168001	AC168001 Bos tauru	C 904	19	57.2	145770	12	AC148384	AC148384 Daeypus n
C 832	19.2	57.8	257703	13	CEV111B2A	AL132904 Caenorhab	C 905	19	57.2	146466	5	AL353637	AL353637 Human DNA
C 833	19.2	57.8	257728	12	AC006846	AC006846 Caenorhab	C 906	19	57.2	147505	6	BS572085	BS572085 Mouse DNA
C 834	19.2	57.8	258174	12	AC079429	AC079429 Mus muscu	C 907	19	57.2	148347	4	AP005408	AP005408 Oryza sat
C 835	19.2	57.8	258274	12	AC105595	AC105595 Rattus no	C 908	19	57.2	148683	11	BS248499	BS248499 Zebrafish
C 836	19.2	57.8	259415	12	AC156178	AC156178 Bos tauru	C 909	19	57.2	148821	6	CT025643	CT025643 Mouse DNA
C 837	19.2	57.8	259679	12	AC169684	AC169684 Bos tauru	C 910	19	57.2	148866	12	AC176538	AC176538 Strongylo
C 838	19.2	57.8	259701	12	AC164709	AC164709 Bos tauru	C 911	19	57.2	149103	12	AC128053	AC128053 Rattus no
C 839	19.2	57.8	259824	12	AC098374	AC098374 Rattus no	C 912	19	57.2	150399	5	AC093680	AC093680 Homo sapi
C 840	19.2	57.8	261771	12	AC125864	AC125864 Rattus no	C 913	19	57.2	151041	4	AP003258	AP003258 Oryza sat
C 841	19.2	57.8	263335	15	CJ11168X5	AL139078 Campyloba	C 914	19	57.2	151345	4	AC135919	AC135919 Oryza sat
C 842	19.2	57.8	279031	12	AC176560	AC176560 Strongylo	C 915	19	57.2	153006	12	AC027782	AC027782 Homo sapi
C 843	19.2	57.8	279504	12	AC169888	AC169888 Bos tauru	C 916	19	57.2	155021	12	AC103385	AC103385 Mus muscu
C 844	19.2	57.8	282690	12	AC164710	AC164710 Bos tauru	C 917	19	57.2	159301	12	CT027752	CT027752 Danio rer
C 845	19.2	57.8	287648	12	AC134140	AC134140 Rattus no	C 918	19	57.2	160599	11	AL954363	AL954363 Zebrafish
C 846	19.2	57.8	289905	12	AC094576	AC094576 Rattus no	C 919	19	57.2	162415	6	AC153986	AC153986 Mus muscu
C 847	19.2	57.8	292390	12	AC105677	AC105677 Rattus no	C 920	19	57.2	162895	12	AC178016	AC178016 Strongylo
C 848	19.2	57.8	293024	12	CEY59A8	Z98870 Caenorhabdi	C 921	19	57.2	163186	5	AC015994	AC015994 Homo sapi
C 849	19.2	57.8	299782	12	AC006844	AC006844 Caenorhab	C 922	19	57.2	163213	12	AC016984	AC016984 Mus muscu
C 850	19.2	57.8	300000	5	AP002534	AP002534 Homo sapi	C 923	19	57.2	164137	12	AC129854	AC129854 Rattus no
C 851	19.2	57.8	322101	2	AX814520	AX814520 Sequence	C 924	19	57.2	165328	6	AC121782	AC121782 Mus muscu
C 852	19.2	57.8	330724	12	CEY67H2	AC022475 Caenorhab	C 925	19	57.2	165799	5	AC1358789	AC1358789 Human DNA
C 853	19.2	57.8	336873	12	AC073666	AC073666 Mus muscu	C 926	19	57.2	167454	12	AC021306	AC021306 Homo sapi
C 854	19.2	57.8	348986	15	BS572093	BS572093 Prochloro	C 927	19	57.2	167546	12	AC163985	AC163985 Rhinolph
C 855	19.2	57.8	349980	2	CQ870293	CQ870293 Sequence	C 928	19	57.2	167862	5	AC055753	AC055753 Homo sapi
C 856	19	57.2	95	15	STWATPROA	M14388 S.lividans	C 929	19	57.2	169147	12	AC129989	AC129989 Rattus no
C 857	19	57.2	184	2	AX905690	AX905690 Sequence	C 930	19	57.2	169547	13	CP000189	CP000189 Drosophil
C 858	19	57.2	184	2	BD041223	BD041223 Sequence	C 931	19	57.2	169742	6	AL646101	AL646101 Mouse DNA
C 859	19	57.2	184	2	AR741484	AR741484 Sequence	C 932	19	57.2	169928	5	AC019067	AC019067 Homo sapi
C 860	19	57.2	366	7	HSB3311XE9	Z67628 H.sapiens D	C 933	19	57.2	170072	6	AC116329	AC116329 Mus muscu
C 861	19	57.2	580	7	BV043651	BV043651 S12P6568	C 934	19	57.2	172697	12	AC076965	AC076965 Homo sapi
C 862	19	57.2	1531	13	AY278950	AY278950 Branchios	C 935	19	57.2	173446	12	AC119178	AC119178 Mus muscu
C 863	19	57.2	2413	5	AK095951	AK095951 Homo sapi	C 936	19	57.2	173474	6	AC127686	AC127686 Mus muscu
C 864	19	57.2	2771	11	CR523878	CR523878 Gallus ga	C 937	19	57.2	174714	12	AC111112	AC111112 Mus muscu
C 865	19	57.2	3325	5	BC009208	BC009208 Homo sapi	C 938	19	57.2	175580	12	AC171304	AC171304 Bos tauru
C 866	19	57.2	3355	5	BC058496	BC058496 Homo sapi	C 939	19	57.2	175920	12	AC164351	AC164351 Bos tauru
C 867	19	57.2	3915	2	AX250162	AX250162 Sequence	C 940	19	57.2	177255	12	AC123235	AC123235 Rattus no
C 868	19	57.2	4136	2	AX455340	AX455340 Sequence	C 941	19	57.2	177800	6	AC157095	AC157095 Mus muscu
C 869	19	57.2	4151	2	CQ492548	CQ492548 Sequence	C 942	19	57.2	177897	11	AC149879	AC149879 Xenopus t
C 870	19	57.2	4151	2	CQ492548	CQ492548 Sequence	C 943	19	57.2	178013	4	AC136219	AC136219 Oryza sat
C 871	19	57.2	6452	15	AY345225	AY345225 Arsenite-	C 944	19	57.2	180227	6	AC138358	AC138358 Mus muscu
C 872	19	57.2	12275	13	AF063309	AF063309 Caenorhab	C 945	19	57.2	180742	6	CT572998	CT572998 Mouse DNA
C 873	19	57.2	16120	11	AP004807	AP004807 Elops sau	C 946	19	57.2	180892	6	AC125533	AC125533 Mus muscu
C 874	19	57.2	16713	11	AB051070	AB051070 Elops haw	C 947	19	57.2	181127	6	AL672038	AL672038 Mouse DNA
C 875	19	57.2	42551	12	CT005259_7	Continuation (8 of	C 948	19	57.2	181544	12	AC167908	AC167908 Bos tauru
C 876	19	57.2	45800	11	AC147865	AC147865 Gopherus	C 949	19	57.2	182189	12	AC118424	AC118424 Rattus no
C 877	19	57.2	48768	5	AL157933	AL157933 Human DNA	C 950	19	57.2	182579	12	AC167513	AC167513 Rattus no
C 878	19	57.2	56013	6	AL773581	AL773581 Mouse DNA	C 951	19	57.2	182627	5	AC096725	AC096725 Homo sapi
C 879	19	57.2	66924	4	MIATGENA_3	Continuation (4 of	C 952	19	57.2	183094	6	AC087903	AC087903 Mus muscu
C 880	19	57.2	73986	12	CT030037_3	Continuation (4 of	C 953	19	57.2	185541	6	AL844537	AL844537 Mouse DNA
C 881	19	57.2	91246	12	LMFLCHR18_10	Continuation (11 o	C 954	19	57.2	185786	12	AC079372	AC079372 Mus muscu
C 882	19	57.2	99878	12	AC087108	AC087108 Homo sapi	C 955	19	57.2	186585	12	AC161803	AC161803 Mus muscu
C 883	19	57.2	100074	6	AL732548	AL732548 Mouse DNA	C 956	19	57.2	190254	11	BS511223	BS511223 Zebrafish
C 884	19	57.2	108457	5	HSJ800J21	AL109955 Human DNA	C 957	19	57.2	190469	6	AC109172	AC109172 Mus muscu
C 885	19	57.2	110000	4	AP008207_139	Continuation (140	C 958	19	57.2	191685	6	AC131663	AC131663 Mus muscu
C 886	19	57.2	110000	4	AP008208_142	Continuation (143	C 959	19	57.2	193164	12	AC179142	AC179142 Strongylo
C 887	19	57.2	110000	4	AP008208_143	Continuation (144	C 960	19	57.2	193304	12	AC109370	AC109370 Rattus no
C 888	19	57.2	110000	4	AP008211_118	Continuation (119	C 961	19	57.2	193386	6	AC116738	AC116738 Mus muscu
C 889	19	57.2	110000	4	AP008211_119	Continuation (120	C 962	19	57.2	194437	12	AC117948	AC117948 Homo sapi
C 890	19	57.2	110000	4	AP008211_185	Continuation (186	C 963	19	57.2	196798	6	AC163035	AC163035 Mus muscu
C 891	19	57.2	110000	12	CT005430_0	AC099430 Rattus no	C 964	19	57.2	199287	6	AC125043	AC125043 Bos tauru
C 892	19	57.2	110000	12	CT005259_6	Continuation (7 of	C 965	19	57.2	199987	12	AC171512	AC171512 Bos tauru
C 893	19	57.2	110000	12	LMFLCHR18_09	Continuation (10 o	C 966	19	57.2	201304	6	AL929012	AL929012 Mouse DNA
C 894	19	57.2	110000	15	AE000513_08	Continuation (9 of	C 967	19	57.2	201748	12	AC161914	AC161914 Bos tauru

```

968 19 57.2 202023 12 AC141270 Homo sapi
c 969 19 57.2 202333 12 AC165202 Oryctolag
970 19 57.2 203604 6 AC158963 Mus muscu
c 971 19 57.2 204206 12 AC176020 Strongylo
972 19 57.2 205101 6 AL645564 Mouse DNA
973 19 57.2 206187 12 AC021172 Homo sapi
c 974 19 57.2 207223 6 AL732521 Mouse DNA
975 19 57.2 207450 12 AC133253 Rattus no
c 976 19 57.2 207782 4 AP003764 Oryza sat
c 977 19 57.2 208140 6 AC131084 Mus muscu
c 978 19 57.2 209409 12 AC164361 Bos tauru
c 979 19 57.2 209885 6 AC153512 Mus muscu
c 980 19 57.2 210190 6 AC15781 Mus muscu
c 981 19 57.2 211056 6 AL591207 Mouse DNA
c 982 19 57.2 211061 6 AC153591 Mus muscu
c 983 19 57.2 211190 12 AC152229 Bos tauru
c 984 19 57.2 211776 12 AC123347 Rattus no
c 985 19 57.2 212081 11 BX005083 Zebrafish
c 986 19 57.2 213783 12 AC098119 Rattus no
c 987 19 57.2 215113 6 AC105167 Mus muscu
c 988 19 57.2 216294 6 AC154470 Mus muscu
c 989 19 57.2 220632 6 AC136517 Mus muscu
c 990 19 57.2 221264 6 AC155258 Mus muscu
c 991 19 57.2 223432 5 AC126755 Homo sapi
c 992 19 57.2 227026 6 AC13476 Mus muscu
c 993 19 57.2 227946 6 AC124175 Mus muscu
c 994 19 57.2 230156 12 AC132773 Rattus no
c 995 19 57.2 230156 12 AC132773 Rattus no
c 996 19 57.2 234584 12 AC111604 Rattus no
c 997 19 57.2 235763 12 AC096378 Rattus no
c 998 19 57.2 238247 12 AC095452 Rattus no
c 999 19 57.2 240349 12 AC096133 Rattus no
c 1000 19 57.2 242228 12 AC135821 Rattus no
19 57.2 243676 6 AC115763 Mus muscu

ALIGNMENTS

RESULT 1
BD233463 603 bp DNA linear PAT 17-JUL-2003
LOCUS Human protein having hydrophobic domain and DNA encoding the same.
DEFINITION BD233463
ACCESSION BD233463
VERSION JP 2002519016-A/9.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Kato,S. and Kimura,T.
AUTHORS Kato,S. and Kimura,T.
TITLE Human protein having hydrophobic domain and DNA encoding the same
JOURNAL Patent: JP 2002519016-A 9 02-JUL-2002;
SAGAMI CHEMICAL RESEARCH CENTER,PROTEGENE INC
COMMENT OS Homo sapiens (human)
PN JP 2002519016-A/9
PD 02-JUL-2002
PF 18-JUN-1999 JP 2000557267
PI SEISHI KATO,TOMOKO KIMURA
PC C12N15/09,C07K14/47,C12N1/15,C12N1/19,C12N5/10,C12N5/10,00,C12N5/ PC
00
CC Human protein having hydrophobic domain and DNA encoding the
FH Key same Location/Qualifiers
FT source 1...603
/organism='Homo sapiens (human)'.
/loc_type="genomic DNA"
/db_xref="taxon:9606"

FEATURES
source
1...603
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

Query Match 90.4%; Score 30; DB 2; Length 603;
Best Local Similarity 64.3%; Pred. No. 0.0065;
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 GNCAYGNTGGCCNTCCNTAYARGAYTAYGARGTNAAR 42
Db 550 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 591
|||:|||||:|||||:|||||:|||||:|||||:|||||:
|||:|||||:|||||:|||||:|||||:|||||:

RESULT 2
BD209699 1212 bp DNA linear PAT 17-JUL-2003
LOCUS Compositions isolated from skin cells and methods for their use.
DEFINITION BD209699
ACCESSION BD209699
VERSION BD209699, 1 GI:33019469
KEYWORDS JP 2002512798-A/171.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.
1 (bases 1 to 1212)
Strachan,L., Sleeman,M., Watson,J.D., Onrust,R., Kumble,A. and
Murison,J.G.
Compositions isolated from skin cells and methods for their use
Patent: JP 2002512798-A 171 08-MAY-2002;
GENESIS RESEARCH AND DEVELOPMENT CORP LTD
COMMENT OS Homo sapiens (human)
PN JP 2002512798-A/171
PD 08-MAY-2002
PF 29-APR-1999 JP 2000546009
PR 29-APR-1998 US 09/069726,09-NOV-1998 US 09/188930 PI
LORNA STRACHAN,MATTHEW SLEEMAN,JAMES DOUGLAS WATSON,RENE PI
ONRUST,
PI ANAND KUMBLE,JAMES GREG MURISON
PC C12N15/09,A61K38/00,A61P9/00,A61P17/00,A61P29/00,A61P31/18, PC
A61P35/00,
C07K14/47,C12N1/15,C12N1/19,C12N1/21,C12N5/10,C12P21/02,C12N15/ PC
00, A61K37/02,C12N5/00
CC Compositions isolated from skin cells and methods for their
FH Key use. Location/Qualifiers
FT source 1..1212
/organism='Homo sapiens (human)'.
/loc_type="genomic DNA"
/db_xref="taxon:9606"

FEATURES
source
1..1212
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 90.4%; Score 30; DB 2; Length 1212;
Best Local Similarity 64.3%; Pred. No. 0.0074;
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 GNCAYGNTGGCCNTCCNTAYARGAYTAYGARGTNAAR 42
Db 990 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1031
|||:|||||:|||||:|||||:|||||:|||||:|||||:
|||:|||||:|||||:|||||:|||||:|||||:

RESULT 3
AR341505 1212 bp DNA linear PAT 17-AUG-2003
LOCUS AR341505
DEFINITION Sequence 249 from patent US 6573095.
ACCESSION AR341505
VERSION AR341505.1 GI:33733640
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unknown.
```

Unclassified.  
REFERENCE 1 (bases 1 to 1212)  
AUTHORS Strachan,L.  
TITLE Polynucleotides isolated from skin cells  
JOURNAL Patent: US 6573095-A 249 03-JUN-2003;  
Genesis Research & Development Corporation Limited; Parnell;  
NZX;  
FEATURES  
source Location/Qualifiers  
1..1212  
/mol\_type="genomic DNA"  
ORIGIN  
Query Match 90.4%; Score 30; DB 2; Length 1212;  
Best Local Similarity 64.3%; Pred. No. 0.0074;  
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;  
Qy 1 GCNCAYGNTGGCCNTCCNTAYAARGAYTAYGAGTNAAR 42  
|||:|||||:|||||:|||||:|||||:|||||:|||||:  
Db 990 GCCCAGCTTGGCGGAGCCCTTACAAAGACTATGAAGTAAAG 1031  
RESULT 4  
AX078375  
LOCUS 1619 bp DNA linear PAT 22-FEB-2001  
DEFINITION Sequence 43 from Patent WO0107612.  
ACCESSION AX078375  
VERSION AX078375.1 GI:13158044  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
REFERENCE 1  
AUTHORS Au-Young,J., Bandman,O., Tang,Y.T., Yue,H., Azimzai,Y., Burford,N.,  
Baughn,M.R., Lu,D.A., Hillman,J.L., Patterson,C. and Lal,P.  
TITLE Receptors and associated proteins  
JOURNAL Patent: WO 0107612-A 43 01-FEB-2001;  
Incyte Genomics, Inc. (US)  
FEATURES  
source Location/Qualifiers  
1..1619  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
/notes="Incyte ID No: 2681738CB1"  
ORIGIN  
Query Match 90.4%; Score 30; DB 2; Length 1619;  
Best Local Similarity 64.3%; Pred. No. 0.0079;  
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;  
Qy 1 GCNCAYGNTGGCCNTCCNTAYAARGAYTAYGAGTNAAR 42  
|||:|||||:|||||:|||||:|||||:|||||:|||||:  
Db 1137 GCCCAGCTTGGCGGAGCCCTTACAAAGACTATGAAGTAAAG 1178  
RESULT 5  
BD233473  
LOCUS 1718 bp DNA linear PAT 17-JUL-2003  
DEFINITION Human protein having hydrophobic domain and DNA encoding the same.  
ACCESSION BD233473  
VERSION BD233473.1 GI:33043243  
KEYWORDS JP 2002519016-A/19.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1718)  
AUTHORS Kato,S. and Kimura,T.  
TITLE Human protein having hydrophobic domain and DNA encoding the same  
JOURNAL Patent: JP 2002519016-A 19 02-JUL-2002;

SAGAMI CHEMICAL RESEARCH CENTER, PROTEGENE INC  
OS Homo sapiens (human)  
PN JP 2002519016-A/19  
PD 02-JUL-2002  
PF 18-JUN-1999 JP 2000557267  
PI SEISHI KATO,TOMOKO KIMURA  
PC C12N15/09,C07K14/47,C12N1/15,C12N1/19,C12N5/10,C12N5/00,C12N5/PC  
00  
CC Human protein having hydrophobic domain and DNA encoding the  
same  
FH Key same  
FT source 1..1718  
Location/Qualifiers  
/organism="Homo sapiens (human)"  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
FEATURES  
source Location/Qualifiers  
1..1718  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Query Match 90.4%; Score 30; DB 2; Length 1718;  
Best Local Similarity 64.3%; Pred. No. 0.0079;  
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;  
Qy 1 GCNCAYGNTGGCCNTCCNTAYAARGAYTAYGAGTNAAR 42  
|||:|||||:|||||:|||||:|||||:|||||:|||||:  
Db 561 GCCCAGCTTGGCGGAGCCCTTACAAAGACTATGAAGTAAAG 602  
RESULT 6  
AK172760  
LOCUS 1826 bp mRNA linear PRI 07-MAY-2004  
DEFINITION Homo sapiens cDNA FLJ23921 fis, clone COL02043, highly similar to  
Homo sapiens retinoic acid induced 3 (RAI3).  
ACCESSION AK172760  
VERSION AK172760.1 GI:47077732  
KEYWORDS oligo capping; fis (full insert sequence).  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
REFERENCE 1  
AUTHORS Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,  
Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,  
Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.  
TITLE NEDO human cDNA sequencing project  
REFERENCE 2 (bases 1 to 1826)  
AUTHORS Sugano,S. and Suzuki,Y.  
JOURNAL Direct Submission  
SUBMITTED (22-APR-2004) Sumio Sugano, Institute of Medical Science,  
University of Tokyo, Laboratory of Genome Structure, Human Genome  
Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan  
(E-mail:flcdna@mail.ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,  
Fax:81-3-5449-5416)  
COMMENT NEDO human cDNA sequencing project supported by Ministry of  
Economy, Trade and Industry of Japan; cDNA full insert sequencing;  
Research Association for Biotechnology; cDNA library construction;  
5'- & 3'-end one pass sequencing; Department of Virology and Human  
Genome Center, Institute of Medical Science, University of Tokyo  
(partly supported by Science and Technology Agency).  
FEATURES  
source Location/Qualifiers  
1..1826  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="COL02043"  
/tissue\_type="colon"  
/clone\_lib="COL"  
/notes="cloning vector pME18SFL3"  
ORIGIN

Query Match Best Local Similarity 90.4%; Score 30; DB 5; Length 1826; Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;			TITLE JOURNAL PUBMED REFERENCE AUTHORS TITLE JOURNAL					
Qy	1	GCNCAYGCGTGGCCNTCNCNTAYARGAYTAYGARGTNAAR 42						
Db	638	GCCACGCTTGGCCGAGCCCTTACAAGACTATGTAAGTAAAG 679						
RESULT 7								
LOCUS	CQ723177	2290 bp	DNA	linear	PAT 03-FEB-2004			
DEFINITION	Sequence 9111 from Patent WO02068579.							
ACCESSION	CQ723177							
VERSION	CQ723177.1							
KEYWORDS	GI:42284034							
SOURCE	Homo sapiens (human)							
ORGANISM	Homo sapiens							
REFERENCE	1							
AUTHORS	Venter,C.J., Adams,M.C., Li,P.W. and Myers,E.W.							
TITLE	Kits, such as nucleic acid arrays, comprising a majority of humanexons or transcripts, for detecting expression and other uses thereof							
JOURNAL	Patent: WO 02068579-A 9111 06-SEP-2002;							
FEATURES	PE Corporation (NV) (US)							
source	Location/Qualifiers							
ORIGIN	1. .2290							
	/organism="Homo sapiens"							
Query Match	90.4%;	Score 30;	DB 2;	Length 2290;				
Best Local Similarity	64.3%;	Pred. No. 0.0084;						
Matches	27;	Conservative	7;	Mismatches 8;	Indels 0; Gaps 0;			
Qy	1	GCNCAYGCGTGGCCNTCNCNTAYARGAYTAYGARGTNAAR 42						
Db	1117	GCCACGCTTGGCCGAGCCCTTACAAGACTATGTAAGTAAAG 1158						
RESULT 8								
LOCUS	BC003665	2296 bp	mRNA	linear	PRI 24-NOV-2004			
DEFINITION	Homo sapiens G protein-coupled receptor, family C, group 5, member A, mRNA (cDNA clone MGC:923 IMAGE:2988011), complete cds.							
ACCESSION	BC003665							
VERSION	BC003665.2							
KEYWORDS	GI:33872669							
SOURCE	Homo sapiens (human)							
ORGANISM	Homo sapiens							
REFERENCE	1							
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.							
Query Match	90.4%;	Score 30;	DB 2;	Length 2290;				
Best Local Similarity	64.3%;	Pred. No. 0.0084;						
Matches	27;	Conservative	7;	Mismatches 8;	Indels 0; Gaps 0;			
Qy	1	GCNCAYGCGTGGCCNTCNCNTAYARGAYTAYGARGTNAAR 42						
Db	1117	GCCACGCTTGGCCGAGCCCTTACAAGACTATGTAAGTAAAG 1158						
RESULT 8								
LOCUS	BC003665	2296 bp	mRNA	linear	PRI 24-NOV-2004			
DEFINITION	Homo sapiens G protein-coupled receptor, family C, group 5, member A, mRNA (cDNA clone MGC:923 IMAGE:2988011), complete cds.							
ACCESSION	BC003665							
VERSION	BC003665.2							
KEYWORDS	GI:33872669							
SOURCE	Homo sapiens (human)							
ORGANISM	Homo sapiens							
REFERENCE	1							
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.							
Query Match	90.4%;	Score 30;	DB 5;	Length 2296;				
Best Local Similarity	64.3%;	Pred. No. 0.0084;						
Matches	27;	Conservative	7;	Mismatches 8;	Indels 0; Gaps 0;			
Qy	1	GCNCAYGCGTGGCCNTCNCNTAYARGAYTAYGARGTNAAR 42						
Db	1112	GCCACGCTTGGCCGAGCCCTTACAAGACTATGTAAGTAAAG 1153						
Query Match	90.4%;	Score 30;	DB 5;	Length 2296;				
Best Local Similarity	64.3%;	Pred. No. 0.0084;						
Matches	27;	Conservative	7;	Mismatches 8;	Indels 0; Gaps 0;			
Qy	1	GCNCAYGCGTGGCCNTCNCNTAYARGAYTAYGARGTNAAR 42						
Db	1112	GCCACGCTTGGCCGAGCCCTTACAAGACTATGTAAGTAAAG 1153						
Query Match	90.4%;	Score 30;	DB 5;	Length 2296;				
Best Local Similarity	64.3%;	Pred. No. 0.0084;						
Matches	27;	Conservative	7;	Mismatches 8;	Indels 0; Gaps 0;			
Qy	1	GCNCAYGCGTGGCCNTCNCNTAYARGAYTAYGARGTNAAR 42						
Db	1112	GCCACGCTTGGCCGAGCCCTTACAAGACTATGTAAGTAAAG 1153						
Query Match	90.4%;	Score 30;	DB 5;	Length 2296;				
Best Local Similarity	64.3%;	Pred. No. 0.0084;						
Matches	27;	Conservative	7;	Mismatches 8;	Indels 0; Gaps 0;			
Qy	1	GCNCAYGCGTGGCCNTCNCNTAYARGAYTAYGARGTNAAR 42						
Db	1112	GCCACGCTTGGCCGAGCCCTTACAAGACTATGTAAGTAAAG 1153						
Query Match	90.4%;	Score 30;	DB 5;	Length 2296;				
Best Local Similarity	64.3%;	Pred. No. 0.0084;						
Matches	27;	Conservative	7;	Mismatches 8;	Indels 0; Gaps 0;			
Qy	1	GCNCAYGCGTGGCCNTCNCNTAYARGAYTAYGARGTNAAR 42						
Db	1112	GCCACGCTTGGCCGAGCCCTTACAAGACTATGTAAGTAAAG 1153						
Query Match	90.4%;	Score 30;	DB 5;	Length 2296;				
Best Local Similarity	64.3%;	Pred. No. 0.0084;						
Matches	27;	Conservative	7;	Mismatches 8;	Indels 0; Gaps 0;			
Qy	1	GCNCAYGCGTGGCCNTCNCNTAYARGAYTAYGARGTNAAR 42						
Db	1112	GCCACGCTTGGCCGAGCCCTTACAAGACTATGTAAGTAAAG 1153						
Query Match	90.4%;	Score 30;	DB 5;	Length 2296;				
Best Local Similarity	64.3%;	Pred. No. 0.0084;						
Matches	27;	Conservative	7;	Mismatches 8;	Indels 0; Gaps 0;			
Qy	1	GCNCAYGCGTGGCCNTCNCNTAYARGAYTAYGARGTNAAR 42						
Db	1112	GCCACGCTTGGCCGAGCCCTTACAAGACTATGTAAGTAAAG 1153						
Query Match	90.4%;	Score 30;	DB 5;	Length 2296;				
Best Local Similarity	64.3%;	Pred. No. 0.0084;						
Matches	27;	Conservative	7;	Mismatches 8;	Indels 0; Gaps 0;			
Qy	1	GCNCAYGCGTGGCCNTCNCNTAYARGAYTAYGARGTNAAR 42						
Db	1112	GCCACGCTTGGCCGAGCCCTTACAAGACTATGTAAGTAAAG 1153						
Query Match	90.4%;	Score 30;	DB 5;	Length 2296;				
Best Local Similarity	64.3%;	Pred. No. 0.0084;						
Matches	27;	Conservative	7;	Mismatches 8;	Indels 0; Gaps 0;			
Qy	1	GCNCAYGCGTGGCCNTCNCNTAYARGAYTAYGARGTNAAR 42						
Db	1112	GCCACGCTTGGCCGAGCCCTTACAAGACTATGTAAGTAAAG 1153						
Query Match	90.4%;	Score 30;	DB 5;	Length 2296;				
Best Local Similarity	64.3%;	Pred. No. 0.0084;						
Matches	27;	Conservative	7;	Mismatches 8;	Indels 0; Gaps 0;			
Qy	1	GCNCAYGCGTGGCCNTCNCNTAYARGAYTAYGARGTNAAR 42						
Db	1112	GCCACGCTTGGCCGAGCCCTTACAAGACTATGTAAGTAAAG 1153						
Query Match	90.4%;	Score 30;	DB 5;	Length 2296;				
Best Local Similarity	64.3%;	Pred. No. 0.0084;						
Matches	27;	Conservative	7;	Mismatches 8;	Indels 0; Gaps 0;			
Qy	1	GCNCAYGCGTGGCCNTCNCNTAYARGAYTAYGARGTNAAR 42						
Db	1112	GCCACGCTTGGCCGAGCCCTTACAAGACTATGTAAGTAAAG 1153						
Query Match	90.4%;	Score 30;	DB 5;	Length 2296;				
Best Local Similarity	64.3%;	Pred. No. 0.0084;						
Matches	27;	Conservative	7;	Mismatches 8;	Indels 0; Gaps 0;			
Qy	1	GCNCAYGCGTGGCCNTCNCNTAYARGAYTAYGARGTNAAR 42						
Db	1112	GCCACGCTTGGCCGAGCCCTTACAAGACTATGTAAGTAAAG 1153						
Query Match	90.4%;	Score 30;	DB 5;	Length 2296;				
Best Local Similarity	64.3%;	Pred. No. 0.0084;						
Matches	27;	Conservative	7;	Mismatches 8;	Indels 0; Gaps 0;			
Qy	1	GCNCAYGCGTGGCCNTCNCNTAYARGAYTAYGARGTNAAR 42						
Db	1112	GCCACGCTTGGCCGAGCCCTTACAAGACTATGTAAGTAAAG 1153						
Query Match	90.4%;	Score 30;	DB 5;	Length 2296;				
Best Local Similarity	64.3%;	Pred. No. 0.0084;						
Matches	27;	Conservative	7;	Mismatches 8;	Indels 0; Gaps 0;			
Qy	1	GCNCAYGCGTGGCCNTCNCNTAYARGAYTAYGARGTNAAR 42						
Db	1112	GCCACGCTTGGCCGAGCCCTTACAAGACTATGTAAGTAAAG 1153						
Query Match	90.4%;	Score 30;	DB 5;	Length 2296;				
Best Local Similarity	64.3%;	Pred. No. 0.0084;						
Matches	27;	Conservative	7;	Mismatches 8;	Indels 0; Gaps 0;			
Qy	1	GCNCAYGCGTGGCCNTCNCNTAYARGAYTAYGARGTNAAR 42						
Db	1112	GCCACGCTTGGCCGAGCCCTTACAAGACTATGTAAGTAAAG 1153						
Query Match	90.4%;	Score 30;	DB 5;	Length 2296;				
Best Local Similarity	64.3%;	Pred. No. 0.0084;						
Matches	27;	Conservative	7;	Mismatches 8;	Indels 0; Gaps 0;			
Qy	1	GCNCAYGCGTGGCCNTCNCNTAYARGAYTAYGARGTNAAR 42						
Db	1112	GCCACGCTTGGCCGAGCCCTTACAAGACTATGTAAGTAAAG 1153						
Query Match	90.4%;	Score 30;	DB 5;	Length 2296;				
Best Local Similarity	64.3%;	Pred. No. 0.0084;						
Matches	27;	Conservative	7;	Mismatches 8;	Indels 0; Gaps 0;			
Qy	1	GCNCAYGCGTGGCCNTCNCNTAYARGAYTAYGARGTNAAR 42						
Db	1112	GCCACGCTTGGCCGAGCCCTTACAAGACTATGTAAGTAAAG 1153						
Query Match	90.4%;	Score 30;	DB 5;	Length 2296;				
Best Local Similarity	64.3%;	Pred. No. 0.0084;						
Matches	27;	Conservative	7;	Mismatches 8;	Indels 0; Gaps 0;			
Qy	1	GCNCAYGCGTGGCCNTCNCNTAYARGAYTAYGARGTNAAR 42						
Db	1112	GCCACGCTTGGCCGAGCCCTTACAAGACTATGTAAGTAAAG 1153						
Query Match	90.4%;	Score 30;	DB 5;	Length 2296;				
Best Local Similarity	64.3%;	Pred. No. 0.0084;						
Matches	27;	Conservative	7;	Mismatches 8;	Indels 0; Gaps 0;			
Qy	1	GCNCAYGCGTGGCCNTCNCNTAYARGAYTAYGARGTNAAR 42						
Db	1112	GCCACGCTTGGCCGAGCCCTTACAAGACTATGTAAGTAAAG 1153						
Query Match	90.4%;	Score 30;	DB 5;	Length 2296;				
Best Local Similarity	64.3%;	Pred. No. 0.0084;						
Matches	27;	Conservative	7;	Mismatches 8;	Indels 0; Gaps 0;			
Qy	1	GCNCAYGCGTGGCCNTCNCNTAYARGAYTAYGARGTNAAR 42						
Db	1112	GCCACGCTTGGCCGAGCCCTTACAAGACTATGTAAGTAAAG 1153						
Query Match	90.4%;	Score 30;	DB 5;	Length 2296;				
Best Local Similarity	64.3%;	Pred. No. 0.0084;						
Matches	27;	Conservative	7;	Mismatches 8;	Indels 0; Gaps 0;			
Qy	1	GCNCAYGCGTGGCCNTCNCNTAYARGAYTAYGARGTNAAR 42						
Db	1112	GCCACGCTTGGCCGAGCCCTTACAAGACTATGTAAGTAAAG 1153						
Query Match	90.4%;	Score 30;	DB 5;	Length 2296;				
Best Local Similarity	64.3%;	Pred. No. 0.0084;						
Matches	27;	Conservative	7;	Mismatches 8;	Indels 0; Gaps 0;			
Qy	1	GCNCAYGCGTGGCCNTCNCNTAYARGAYTAYGARGTNAAR 42						
Db	1112	GCCACGCTTGGCCGAGCCCTTACAAGACTATGTAAGTAAAG 1153						
Query Match	90.4%;	Score 30;	DB 5;	Length 2296;				
Best Local Similarity	64.3%;	Pred. No. 0.0084;						
Matches	27;	Conservative	7;	Mism				

```
RESULT 9
LOCUS AF506289 2297 bp mRNA linear PRI 15-JUL-2002
DEFINITION Homo sapiens orphan G protein-coupling receptor PBIG-1 mRNA,
complete cds.
ACCESSION AF506289
VERSION AF506289.1 GI:21779962
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1 (bases 1 to 2040)
Cafferata,E.G., Gonzalez-Guerrico,A.M., Pivetta,O.H. and
Santa-Coloma,T.A.
Identification by differential display of a mRNA specifically
induced by 12-O-tetradecanoylphorbol-13-acetate (TPA) in T84 human
colon carcinoma cells
Cell. Mol. Biol. 42 (5), 797-804 (1996)
JOURNAL
PUBMED 8832110
REFERENCE
2 (bases 1 to 2297)
Cafferata,E.G., Gonzalez-Guerrico,A.M., Costanzo,R., Pivetta,O.H.
and Santa-Coloma,T.A.
Direct Submission
Submitted (24-APR-2002) Laboratorio de Biologia Celular y
Molecular, Instituto de Investigaciones Bioquimicas Fundacion
Campomar, Patricia Argentina 435, Buenos Aires 1405, Argentina
Campomar,
Location/Qualifiers
FEATURES
source
1 .2297
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="12"
/map="12p13-p12.3"
/sex="male"
/cell_line="T84"
/tissue_type="colorectal carcinoma"
/notes="derived from dbEST BE519991"
108..1181
/feature="belongs to the type 3 G protein-coupling receptor
family; characterized by a 7-transmembrane domain motif;
phorbol ester induced protein-1"
/codon_start=1
/product="orphan G protein-coupling receptor PBIG-1"
/protein_id="AAAM77594.1"
/db_xref="GI:21779963"
/translation="MATTVPDGCNGLSKYRLCDKAEAWGIVLETATAGVVTVA
FMULTPLVCKVQDSNRKMLPTQFLFLGLVGLFTFAFIIGLDGSGTPRFLFG
ILFSICSLAHAVSLTKLVGRKPLSLVILGLAVGFLSLQDVIAIEYILTMRT
NNVFSLSAPRRNEDFVLLTVYVFLMALTFMLSSFTFCGSGTGWKRHGAHYLTML
LSIAIWTATLMLPDRDDTILSSAALANGWVFLLAYVSPEFWLTGKRNPM
YPVEDAFKPLVKKGVENRAYSQEITQGFETGTLYPYSTHFLQNPQPOKE
FSIPRAHAFSPYKDYEVKKEGS"
ORIGIN
Query Match 90.4%; Score 30; DB 5; Length 2297;
Best Local Similarity 64.3%; Pred. No. 0.0084;
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;
Qy 1 GCNCAYCGTGGCCNTCNCNTAYARGAYTAGTGTTAAR 42
|||||
Db 1125 GCCACGCTGGCGCGCCCTTACAAAGACTATGAGTAAAG 1166
|||||
RESULT 10
LOCUS AX930411 2302 bp DNA linear PAT 22-DEC-2003
DEFINITION Sequence 2 from Patent WO03087832.
ACCESSION AX930411
VERSION AX930411.1 GI:40312209
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1 (bases 1 to 2040)
Cafferata,E.G., Gonzalez-Guerrico,A.M., Pivetta,O.H. and
Santa-Coloma,T.A.
Identification by differential display of a mRNA specifically
induced by 12-O-tetradecanoylphorbol-13-acetate (TPA) in T84 human
colon carcinoma cells
Cell. Mol. Biol. 42 (5), 797-804 (1996)
JOURNAL
PUBMED 8832110
REFERENCE
2 (bases 1 to 2297)
Cafferata,E.G., Gonzalez-Guerrico,A.M., Costanzo,R., Pivetta,O.H.
and Santa-Coloma,T.A.
Direct Submission
Submitted (24-APR-2002) Laboratorio de Biologia Celular y
Molecular, Instituto de Investigaciones Bioquimicas Fundacion
Campomar, Patricia Argentina 435, Buenos Aires 1405, Argentina
Campomar,
Location/Qualifiers
FEATURES
source
1 .2297
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="12"
/map="12p13-p12.3"
/sex="male"
/cell_line="T84"
/tissue_type="colorectal carcinoma"
/notes="derived from dbEST BE519991"
108..1181
/feature="belongs to the type 3 G protein-coupling receptor
family; characterized by a 7-transmembrane domain motif;
phorbol ester induced protein-1"
/codon_start=1
/product="orphan G protein-coupling receptor PBIG-1"
/protein_id="AAAM77594.1"
/db_xref="GI:21779963"
/translation="MATTVPDGCNGLSKYRLCDKAEAWGIVLETATAGVVTVA
FMULTPLVCKVQDSNRKMLPTQFLFLGLVGLFTFAFIIGLDGSGTPRFLFG
ILFSICSLAHAVSLTKLVGRKPLSLVILGLAVGFLSLQDVIAIEYILTMRT
NNVFSLSAPRRNEDFVLLTVYVFLMALTFMLSSFTFCGSGTGWKRHGAHYLTML
LSIAIWTATLMLPDRDDTILSSAALANGWVFLLAYVSPEFWLTGKRNPM
YPVEDAFKPLVKKGVENRAYSQEITQGFETGTLYPYSTHFLQNPQPOKE
FSIPRAHAFSPYKDYEVKKEGS"
ORIGIN
Query Match 90.4%; Score 30; DB 5; Length 2297;
Best Local Similarity 64.3%; Pred. No. 0.0084;
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;
Qy 1 GCNCAYCGTGGCCNTCNCNTAYARGAYTAGTGTTAAR 42
|||||
Db 1125 GCCACGCTGGCGCGCCCTTACAAAGACTATGAGTAAAG 1166
|||||
RESULT 11
LOCUS AF095448 2302 bp mRNA linear PRI 29-DEC-1998
DEFINITION Homo sapiens putative G protein-coupled receptor (RAIG1) mRNA,
complete cds.
ACCESSION AF095448
VERSION AF095448.1 GI:4063889
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1 (bases 1 to 2302)
Cheng,Y. and Lotan,R.
Molecular cloning and characterization of a novel retinoic
acid-inducible gene that encodes a putative G protein-coupled
receptor
J. Biol. Chem. 273 (52), 35008-35015 (1998)
JOURNAL
PUBMED 9857033
REFERENCE
2 (bases 1 to 2302)
Cheng,Y. and Lotan,R.
Direct Submission
Submitted (27-SEP-1998) Tumor Biology, The University of Texas M.
D. Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX
77030, USA
Location/Qualifiers
FEATURES
source
1 .2302
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="12"
/map="12p12.3-p13"
1..2302
/genes="RAIG1"
/notes="induced by retinoic acid"
100..1173
/codon_start=1
/gene="RAIG1"
/product="putative G protein-coupled receptor"
/protein_id="AAC98506.1"
/db_xref="GI:4063890"
/translation="MATTVPDGCNGLSKYRLCDKAEAWGIVLETATAGVVTVA
FMULTPLVCKVQDSNRKMLPTQFLFLGLVGLFTFAFIIGLDGSGTPRFLFG
ILFSICSLAHAVSLTKLVGRKPLSLVILGLAVGFLSLQDVIAIEYILTMRT
NNVFSLSAPRRNEDFVLLTVYVFLMALTFMLSSFTFCGSGTGWKRHGAHYLTML
NNVFSLSAPRRNEDFVLLTVYVFLMALTFMLSSFTFCGSGTGWKRHGAHYLTML"
ORIGIN
Query Match 90.4%; Score 30; DB 2; Length 2302;
Best Local Similarity 64.3%; Pred. No. 0.0084;
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;
Qy 1 GCNCAYCGTGGCCNTCNCNTAYARGAYTAGTGTTAAR 42
|||||
Db 1117 GCCACGCTGGCGCGCCCTTACAAAGACTATGAGTAAAG 1158
|||||
```

LSIAIIVAMITLLMLPDDFRRWDDTILSSALAANGVFLLAYVSPFWLLTKORNPMD  
YPVEDAFCKPQLVKYSYGVENRAYSQEETIQGFEETGDTLYAPYSTHFLQNPQPOKE  
FSIPRAHWPSPYKDYEVKKEGS"

ORIGIN

Query Match 90.4%; Score 30; DB 5; Length 2302;  
Best Local Similarity 64.3%; Pred. No. 0.0084;  
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 GCNCAYGCGTGGCCNTCNCNTAYARGAYTAYGARGTNAAR 42  
|||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
Db 1117 GCCCAGCGCTGGCGAGCCCTTACAAGACTATGAAGTAAAG 1158

RESULT 12

BD156680 2446 bp DNA linear PAT 17-JAN-2003  
LOCUS  
DEFINITION  
Primer for synthesizing full-length cDNA and use thereof.

BD156680

BD156680.1 GI:27862438

JP 2002191363-A/11523

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominiidae; Homo.

1 (bases 1 to 2446)

Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,

Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.

Primer for synthesizing full-length cDNA and use thereof

Patent: JP 2002191363-A 11523 09-JUL-2002;

HELIX RESEARCH INSTITUTE

OS Homo sapiens (human)

PN JP 2002191363-A/11523

PD 09-JUL-2002

PF 28-JUL-2000 JP 2000280990

PI TOSHIO OTA,TAKAO ISOGAI,TETSUO NISHIKAWA,KOJI HAYASHI,KAORU

PI SAITO,

PI JUNCHI YAMAMOTO,SHIZUKO ISHII,TOMOYASU SUGIYAMA,AI WAKAMATSU,

PI KEIICHI NAGAI,TETSUJI OTSUKI

PC

C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC

10,

PC C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC

Primer for synthesizing full-length cDNA and use thereof FH Key

Location/Qualifiers

FT CDS (254)..(1324).

Location/Qualifiers

1..2446

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

ORIGIN

Query Match 90.4%; Score 30; DB 2; Length 2446;  
Best Local Similarity 64.3%; Pred. No. 0.0085;  
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 GCNCAYGCGTGGCCNTCNCNTAYARGAYTAYGARGTNAAR 42  
|||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
Db 1271 GCCCAGCGCTGGCGAGCCCTTACAAGACTATGAAGTAAAG 1312

RESULT 13

AX877483 2446 bp DNA linear PAT 17-DEC-2003  
LOCUS  
DEFINITION  
Sequence 12388 from Patent EP1074617.

AX877483

AX877483.1 GI:40032219

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

CDS

ORIGIN

Query Match

Best Local Similarity

Matches

27; Conservative

7; Mismatches

8; Indels

0; Gaps

0;

QY

Db

1271

GCCCAGCGCTGGCGAGCCCTTACAAGACTATGAAGTAAAG 1312

RESULT 14

AK001761

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominiidae; Homo.

1

Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,

Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.

Primer for synthesizing full-length cDNA and their use

Patent: EP 1074617-A 12388 07-FEB-2001;

Research Association for Biotechnology (JP)

Location/Qualifiers

1..2446

/organism="Homo sapiens"

/mol\_type="unassigned DNA"

/db\_xref="taxon:9606"

254..1327

/note="unnamed protein product"

/codon\_start=1

/protein\_id="CAE89746.1"

/db\_xref="GI:40032220"

/translation="MATVDPGCRNGLKSKYVRLCDKAEAWGIVLETATAGVVTVA

FMLTPILVKVQDSNRRLMPTQFLGLVIGLIGLTFAFIIGLDSGTGPTFFLFG

ILFSICFSCILAHAVSLTKLVRGKPLSLVILGLAVGSLVQDVIAIYVLTMR

NVNVESELAPRNEEDFVLLTVLFLMALTFMSFTFCGSGTGWKRGHAHLYTML

LSIAIIVAMITLLMLPDDFRRWDDTILSSALAANGVFLLAYVSPFWLLTKORNPMD

YPVEDAFCKPQLVKYSYGVENRAYSQEETIQGFEETGDTLYAPYSTHFLQNPQPOKE

FSIPRAHWPSPYKDYEVKKEGS"

Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T., Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K., Nagase, T., Nomura, N., Kikuchi, H., Masuho, Y., Yamashita, R., Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S. Complete sequencing and characterization of 21,243 full-length human cDNAs  
Nat. Genet. 36 (1), 40-45 (2004)  
14702039

2  
Isogai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y., Nishikawa, T., Nagai, K., Sugano, S., Ishibashi, T., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M., Hiraoka, S., Ishii, S., Kawai, Y., Saito, K., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagahara, K., Masuho, Y. and Kanehori, K.  
NEDO human cDNA sequencing project  
Unpublished  
3 (bases 1 to 2446)  
Isogai, T. and Otsuki, T.  
Direct Submission  
Submitted (16-FEB-2000) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: flj-cdna@nifty.com, Tel:81-438-52-3975, Fax:81-438-52-3986)  
NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing and clone selection; Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.  
Location/Qualifiers  
1. .2446  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="NT2RP5003506"  
/cell\_line="NT2"  
/cell\_type="teratocarcinoma"  
/clone\_lib="NT2RP5"  
/notes="cloning vector: pME18SFL3  
mRNA from NT2 neuronal precursor cells after 2-weeks retinoic acid (RA) induction."  
254. .1327  
/notes="unnamed protein product"  
/codon\_start=1  
/protein\_id="BAA91890.1"  
/db\_xref="GI:7023230"  
/translation="MATTVPDGRNGLSKYYRLCDKAEAWGIVLETATAGVVTSA  
FMLTPLIVCKVQDSNRKMLPTQFLFLGLVGLFGLTFAPFIIGLDSTGPTFFLFG  
ILFICFSLCAHVASLTGLVRGKSLILVGLAVGSLVDVIAEIVLTWMT  
NNVFSLSAPRNEPFLVLLTYVLFMLALTFLMSSFTFCGSGTGWKRHGLVLTWL  
LSIAIWNATILMLDFDRRWDITLSSALAANGWVFLAYVSPFWLLTKORNPMD  
YPVEDAFCKPOLVKKSGVSNRAYSOEEITQGFETGDTLYAPSTHYFQLQNPPOKE  
FSIPRAHWSPYKDYEVKKEGS"

CDS  
Query Match 90.4%; Score 30; DB 5; Length 2446;  
Best Local Similarity 64.3%; Pred. No. 0.0085;  
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

ORIGIN  
1 GCNCAYCGNTGGCCNTCCNTAYAAAGATYAGACTGAAGTAAAG 42  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
Db 1271 GCCCAGCGTTGGCGGAGCCCTTACAAAGACTATGAAGTAAAG 1312

RESULT 15  
LOCUS CQ981495 2456 bp DNA linear PAT 25-JAN-2005  
DEFINITION Sequence 350 from Patent EP1498424.  
ACCESSION CQ981495  
VERSION CQ981495.1 GI:58190785  
KEYWORDS Homo sapiens (human)  
SOURCE

Query Match 90.4%; Score 30; DB 5; Length 2446;  
Best Local Similarity 64.3%; Pred. No. 0.0085;  
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

ORIGIN  
1 GCNCAYCGNTGGCCNTCCNTAYAAAGATYAGACTGAAGTAAAG 42  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
Db 1271 GCCCAGCGTTGGCGGAGCCCTTACAAAGACTATGAAGTAAAG 1312

Query Match 90.4%; Score 30; DB 2; Length 2456;  
Best Local Similarity 64.3%; Pred. No. 0.0085;  
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

ORIGIN  
1 GCNCAYCGNTGGCCNTCCNTAYAAAGATYAGACTGAAGTAAAG 42  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
Db 1271 GCCCAGCGTTGGCGGAGCCCTTACAAAGACTATGAAGTAAAG 1312

Query Match 90.4%; Score 30; DB 2; Length 2456;  
Best Local Similarity 64.3%; Pred. No. 0.0085;  
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

ORIGIN  
1 GCNCAYCGNTGGCCNTCCNTAYAAAGATYAGACTGAAGTAAAG 42  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
Db 1271 GCCCAGCGTTGGCGGAGCCCTTACAAAGACTATGAAGTAAAG 1312

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.

REFERENCE  
AUTHORS Rosenthal, A., Hermann, K., Heiden, E., Pilarsky, C., Bruemendorf, T.,  
Staub, E., Roepcke, S., Mennerich, D., Kinnenmann, H. and Li, X.  
Human nucleic acid sequences from lung tumours  
Patent: EP 1498424-A 350 19-JAN-2005;  
Hinzmann, Bernd (DE); Hermann, Klaus (DE); Heiden, Esmeralda (DE);  
Rosenthal, Andre (DE)  
FEATURES  
source 1. .2456  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

ORIGIN  
Query Match 90.4%; Score 30; DB 2; Length 2456;  
Best Local Similarity 64.3%; Pred. No. 0.0085;  
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

Qy 1 GCNCAYCGNTGGCCNTCCNTAYAAAGATYAGACTGAAGTAAAG 42  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
Db 1271 GCCCAGCGTTGGCGGAGCCCTTACAAAGACTATGAAGTAAAG 1312

RESULT 16  
LOCUS DD210040 2456 bp DNA linear PAT 19-JAN-2006  
DEFINITION Methods of Diagnosis of Cancer, Compositions and Methods of  
Screening for Modulators of Cancer.  
ACCESSION DD210040.1 GI:85654022  
KEYWORDS JP 2005518782-A/34.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.

REFERENCE  
AUTHORS Zlotnik, A., Mack, D.H., Agiz, N., Gish, K.C., Hebeji, P.A., Wilson, K.E.  
and Afar, D.  
TITLE Methods of Diagnosis of Cancer, Compositions and Methods of  
Screening for Modulators of Cancer  
Patent: JP 2005518782-A 34 30-JUN-2005;  
JOURNAL PROTEIN DESIGN LABS INC  
COMMENT OS Homo sapiens  
PN JP 2005518782-A/34  
PD 30-JUN-2005  
PF 17-SEP-2002 JP 2003529912  
PR 12-APR-2002 US 60/372246, 08-FEB-2002 US 60/355257, PR  
08-FEB-2002 US 60/355145, 13-NOV-2001 US 60/350666, PR  
20-FEB-2001 US 60/323887, 17-SEP-2001 US 60/323469 PI albert  
zlotnik,david h mack,natasha agiz,kurt c gish,peter a pi hebeji,  
pi keith e wilson,daniel afar  
CC  
FH Key Location/Qualifiers.  
1. .2456  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

FEATURES  
source 1. .2456  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

ORIGIN  
Query Match 90.4%; Score 30; DB 2; Length 2456;  
Best Local Similarity 64.3%; Pred. No. 0.0085;  
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

Qy 1 GCNCAYCGNTGGCCNTCCNTAYAAAGATYAGACTGAAGTAAAG 42  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
Db 1271 GCCCAGCGTTGGCGGAGCCCTTACAAAGACTATGAAGTAAAG 1312





JOURNAL Patent: WO 0142467-A 4043 14-JUN-2001;  
Millennium Predictive Medicine, Inc. (US)  
FEATURES Location/Qualifiers  
source 1..3371  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Query Match 90.4%; Score 30; DB 2; Length 3371;  
Best Local Similarity 64.3%; Pred. No. 0.009;  
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;  
QY 1 GCNCAYGCNTGGCCNTCNCNTAYARGAYTAYGAGTNAAR 42  
Db 1184 GCCCAGCCTTGGCGGACCCCTTACAAAGACTATGAGTAAG 1225  
RESULT 20  
CQ894732  
LOCUS 6730 bp DNA linear PAT 05-NOV-2004  
DEFINITION Sequence 42 from Patent EP1471075.  
ACCESSION CQ894732  
VERSION CQ894732.1 GI:55467481  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
REFERENCE 1  
AUTHORS Rosenthal,A.D., Pilarsky,C., Dahl,E., Specht,T., Bruemmendorf,T.,  
Lichtner,R., Staub,E., Roepcke,S. and Li,X.I.  
TITLE Human nucleic acid sequences expressed in pancreatic carcinomas  
JOURNAL Patent: EP 1471075-A 42 27-OCT-2004;  
Hinzmann, Bernd (DE); Rosenthal, Andre (DE); Pilarsky, Christian  
(DE); Dahl, Edgar (DE); Specht, Thomas (DE); Lichtner, Rosemarie  
(DE)  
FEATURES Location/Qualifiers  
source 1..6730  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Query Match 90.4%; Score 30; DB 2; Length 6730;  
Best Local Similarity 64.3%; Pred. No. 0.01;  
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;  
QY 1 GCNCAYGCNTGGCCNTCNCNTAYARGAYTAYGAGTNAAR 42  
Db 1271 GCCCAGCCTTGGCGGACCCCTTACAAAGACTATGAGTAAG 1312  
RESULT 21  
AC007688/c  
LOCUS 161577 bp DNA linear PRI 30-AUG-2002  
DEFINITION Homo sapiens 12 BAC RPC111-392P7 (Roswell Park Cancer Institute  
Human BAC Library) complete sequence.  
ACCESSION AC007688  
VERSION AC007688.15 GI:5815499  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
REFERENCE 1 (bases 1 to 161577)  
AUTHORS Muzny,D., Arenson,A.D., Bouck,J., Bunac,C., Chen,Z., Ding,Y.,  
Dugan,S., Durbin,J., Forcum,J., Garcia,C., Gorrell,J.H.,  
Gorrell,L.L., Hernandez,J., Issar,A., Jackson,L., Kneitz,S.,  
Kondejewski,N., Lau,S., Leal,B., Lee,E., Lichtarge,O., Liu,W.,  
Logan,O., Lu,J., Marondel,I., Martinez,C., Merscher,S., Miller,A.,  
Montgomery,K., Oswal,G., Pampell,L.R., Parish,B.J., Perez,L.,  
Rashid,N.D., Rives,C., Scherer,S.E., Shen,H., Shim,C., Simon,M.,  
Vo,Q., Williamson,A., Worley,K.C., Xhang,A.M., Yang,R., Yu,W.,  
Zhou,X., Kucherlapati,R., Nelson,D. and Gibbs,R.A.  
Direct Submission  
Unpublished  
2 (bases 1 to 161577)  
Worley,K.C.  
Direct Submission  
Submitted (01-JUN-1999) Molecular and Human Genetics, Baylor  
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 161577)  
Worley,K.C.  
Direct Submission  
Submitted (01-SEP-1999) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
4 (bases 1 to 161577)  
Worley,K.C.  
Direct Submission  
Submitted (16-MAY-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
5 (bases 1 to 161577)  
Worley,K.C.  
Direct Submission  
Submitted (30-AUG-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Sep 1, 1999 this sequence version replaced gi:5757565.  
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email  
[gc-help@bcm.tmc.edu](mailto:gc-help@bcm.tmc.edu)  
CLONE LENGTH: This sequence does not necessarily represent the  
entire insert of this clone. Overlapping regions of clones are only  
sequenced and submitted once, so the sequence for the remainder of  
the insert may be found in the record for the adjacent clones.  
Overlapping clones are noted at the beginning and end of the  
Features listing.  
ANNOTATION OF FEATURES:  
STSs are identified using ePCR (Genome Res. 7:541-550) searches  
of a local database that includes entries from dbSTS, GDB, and  
local mapping efforts.  
Repeats are identified using RepeatMasker (A. Smit and P. Green,  
unpublished.) for Human and Mouse sequences.  
Genes and region of sequence similarity are identified by BLAST  
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the  
EST and cDNA sequences. Genes demonstrate at least two exons  
flanked by consensus splice sites that maintained sequence  
continuity across the splice junctions. Sequences that are not  
identical matches are annotated as similar.  
SEQUENCING READ COVERAGE: Sequencing is completed to a minimum  
standard of double strand coverage with a minimum of 2 clones and 2  
reads with no ambiguities or 2 chemistries with a minimum of 2  
clones and 3 reads with no ambiguities. If the sequence quality for  
a region does not meet this standard, it will be indicated in the  
annotation as Low Coverage.  
QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality  
standards - estimated error rate less than 1 per 10,000 bases.  
Reports of lowest quality individual bases and measures of base  
quality are listed below. Description of the metrics can be found  
at URL:  
<http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.  
QUALSTAT-REPORT-----  
----- Summary Statistics -----  
Contig length: 161577  
Phrap values in estimate: 160751

Average error rate (BCM-Phrap estimate):		0.000163681
Fraction of Phrap values less than 40 :		0.0376047
Number of consensus changing edits:		30
Number of N's in consensus :		0
----- Consensus changing edits -----		
Position	Original+Context	Edited+Context
7033	acctgcgcgt(n)ccgcgccctt	acctgcgcgt(c)ccgcgccctt
47567	aaaaaaaaa(n)ggaataaat	aaaaaaaaa(a)ggaataaat
51135	aaagaagaa(n)aaagaagaa	aaagaagaa(a)aaagaagaa
75582	aaaaaaaaa(n)angagatgt	aaaaaaaaa(a)aaaggatgt
75585	aaaaaaa(n)ggatgttcgc	aaaaaaaaa(a)ggatgttcgc
75667	ctaaagcaga(n)taagatctta	ctaaagcaga(a)taagatctta
75752	ttttaaatag(n)gctttgcttt	ttttaaatag(t)gctttgcttt
84017	ggaggggaag(n)aaggaaggag	ggaggggaag(g)aaggaaggag
85227	tttgttttt(n)tgttttttt	tttgttttt(t)tgttttttt
99681	ggagttcaag(n)atgcagtgag	ggagttcaag(g)atgcagtgag
111307	aatctcttat(n)ccgaattca	aatctcttat(g)ccgaattca
111374	actagataac(n)atcnttttt	actagataac(t)atcnttttt
111378	gtatacnatc(n)ttttttttt	gtatacnatc(c)ttttttttt
112621	aaaaaaaaa(n)ccatctctaga	aaaaaaaaa(a)ccatctctaga
135812	atcacccctc(c)ttttttttt	atcacccctc(t)ttttttttt
137207	ttgcaggcac(n)cgccaccag	ttgcaggcac(a)cgccaccag
137218	cgcaccacg(n)ctggctaggt	cgcaccacg(c)ctggctaggt
145113	gcaagtgaan(n)natgtagaat	gcaagtgaan(a)natgtagaat
145114	caagtgaan(n)atgtagaata	caagtgaan(g)atgtagaata
145232	ggcacgggtg(n)ntcacnccag	ggcacgggtg(c)ntcacnccag
145233	gcaccgggtn(n)tcacnccagt	gcaccgggtg(c)tcacnccagt
145238	ggtgnntcac(n)ccagtaatcc	ggtgnntcac(g)ccagtaatcc
145321	gcaacatggt(t)nnccccatc	gcaacatggt(g)nnccccatc
145322	caacatggtt(n)nnccccatc	caacatggtg(a)nnccccatc
145323	aacatggttn(n)ccccatctc	aacatggtg(a)ccccatctc
145324	acatggttnn(n)ccccatctt	acatggtgaa(a)ccccatctt
145363	gttcaccagg(n)gtggtggcgt	gttcaccagg(t)gtggtggcgt
145377	gtggcgta(n)tgtgtgacct	gtggcgta(g)tgtgtgacct
145397	tagctacttg(n)gagggaggat	tagctacttg(g)gagggaggat
145505	aaaaaaaaa(n)gaaaaaaa	aaaaaaaaa(a)gaaaaaaa
----- Distribution of Quality < 40 Bases -----		
#	1000	*
	900	*
	800	*
	700	*
	600	*
bases	500	*
	400	*
	300	*
	200	*
	100	*
	0	*
-----		
Version: 1.01 qxfo.		
Location/Qualifiers		
1. .161577		
/organism="Homo sapiens"		
/mol_type="genomic DNA"		
/db_xref="taxon:9606"		
/chromosomes="12"		
/clone="RP11-392p7"		
complement(3...65)		
/rpt_family="MER4B"		
161. .471		
/rpt_family="AluXs"		
complement(607. .702)		
/rpt_family="MIR"		
-----		
repeat_region	1 (bases 1 to 680)	
repeat_region	Ota.T., Isogai.T., Nishikawa.T., Hayashi.K., Saito.K., Yamamoto.J., Ishii.S., Sugiyama.T., Wakamatsu.A., Nagai.K. and Otsuki.T.	
repeat_region	Primer for synthesizing full-length cDNA and use thereof	
repeat_region	Patent: JP 2002191363-A 5189 09-JUL-2002;	
repeat_region	HELIX RESEARCH INSTITUTE	
repeat_region	OS Homo sapiens (human)	
repeat_region	EN JP 2002191363-A/5189	
repeat_region	PD 09-JUL-2002	
repeat_region	PF 28-JUL-2000 JP 2002280990	
repeat_region	PI TOSHIO OTA,TAKAO ISOGAI,TETSUO NISHIKAWA,KOJI HAYASHI,KAORU	
repeat_region	PI SAITO,	
repeat_region	PI JUNICHI YAMAMOTO,SHIZUKO ISHII,TOMOYASU SUGIYAMA,AI WAKAMATSU,	
repeat_region	PI KEIICHI NAGAI,TETSUJI OTSUKI	

repeat_region	703. .838
repeat_region	/rpt_family="FLAM_C"
repeat_region	complement(839. .916)
repeat_region	/rpt_family="MIR"
repeat_region	complement(1205. .1377)
repeat_region	/rpt_family="MER104"
repeat_region	1388. .1525
repeat_region	/standard_name="WIAF-759-STS"
repeat_region	/db_xref="dbSTS:65358"
repeat_region	1683. .1971
repeat_region	/rpt_family="AluJo"
repeat_region	1974. .2261
repeat_region	/rpt_family="AluJb"
repeat_region	2486. .2593
repeat_region	/rpt_family="MIR"
repeat_region	2610. .3214
repeat_region	/note="Region:Regione cluster similar to AA056332 and
repeat_region	AI074576"
repeat_region	2712. .2823
repeat_region	/standard_name="SHGC-44583"
repeat_region	/db_xref="dbSTS:48451"
repeat_region	3185. .3491
repeat_region	/rpt_family="AluSq"
repeat_region	complement(6397. .6537)
repeat_region	/rpt_family="L2"
repeat_region	6967. .7158
repeat_region	/rpt_family="(CCCCG)n"
repeat_region	7094. .7240
repeat_region	/rpt_family="(CCG)n"
repeat_region	complement(7770. .8250)
repeat_region	/rpt_family="MER44B"
repeat_region	8831. .9073
repeat_region	/rpt_family="MER102"
repeat_region	complement(9075. .9303)
repeat_region	/rpt_family="L2"
repeat_region	9671. .10135
repeat_region	/rpt_family="LTR33A"
repeat_region	90.4%; Score 30; DB 5; Length 161577;
repeat_region	Best Local Similarity 64.3%; Pred. No. 0.019;
repeat_region	Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;
repeat_region	QY 1 GCNCAAGTGGCCNTCCNTAYARGAYTAGGTNAAR 42
repeat_region	DB 94911 GCCACGCTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 94870
repeat_region	RESULT 22
repeat_region	BD150346
repeat_region	LOCUS
repeat_region	BD150346
repeat_region	Primer for synthesizing full-length cDNA and use thereof.
repeat_region	DEFINITION
repeat_region	BD150346
repeat_region	ACCESSION
repeat_region	BD150346.1 GI:27856104
repeat_region	VERSION
repeat_region	JP 2002191363-A/5189.
repeat_region	KEYWORDS
repeat_region	Homo sapiens (human)
repeat_region	SOURCE
repeat_region	Homo sapiens
repeat_region	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
repeat_region	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
repeat_region	Hominidae; Homo.
repeat_region	1 (bases 1 to 680)
repeat_region	Ota.T., Isogai.T., Nishikawa.T., Hayashi.K., Saito.K., Yamamoto.J.,
repeat_region	Ishii.S., Sugiyama.T., Wakamatsu.A., Nagai.K. and Otsuki.T.
repeat_region	Primer for synthesizing full-length cDNA and use thereof
repeat_region	Patent: JP 2002191363-A 5189 09-JUL-2002;
repeat_region	HELIX RESEARCH INSTITUTE
repeat_region	OS Homo sapiens (human)
repeat_region	EN JP 2002191363-A/5189
repeat_region	PD 09-JUL-2002
repeat_region	PF 28-JUL-2000 JP 2002280990
repeat_region	PI TOSHIO OTA,TAKAO ISOGAI,TETSUO NISHIKAWA,KOJI HAYASHI,KAORU
repeat_region	PI SAITO,
repeat_region	PI JUNICHI YAMAMOTO,SHIZUKO ISHII,TOMOYASU SUGIYAMA,AI WAKAMATSU,
repeat_region	PI KEIICHI NAGAI,TETSUJI OTSUKI

```
PC
C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC
10, C12N21/02,C12Q1/68/C12P21/08,C06F17/30,C12N15/00,C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof PH Key
Location/Qualifiers
FT source
1. .680
/organism='Homo sapiens (human)'.
FEATURES
source
1..680
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 82.5%; Score 27.4; DB 2; Length 680;
Best Local Similarity 59.5%; Pred. No. 0.13;
Matches 25; Conservative 7; Mismatches 10; Indels 0; Gaps 0;
QY 1 GCNCAYCNGTGGCCNTCNCNTAYAARGAYTAYGAGTNAAR 42
|||:||||| |||||:|||||:|||||: |||
Db 633 GCCCAGCTTGGCCCAACCTTACAAAGACTATGAANGTAA 674
|||:||||| |||||:|||||:|||||: |||
RESULT 23
AX870284 680 bp DNA linear PAT 17-DEC-2003
LOCUS
DEFINITION Sequence 5189 from Patent EP1074617.
ACCESSION AX870284
VERSION AX870284.1 GI:40025147
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
REFERENCE
1 Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
Primers for synthesizing full-length cDNA and their use
Patent: EP 1074617-A 5189 07-FEB-2001;
Research Association for Biotechnology (JP)
Location/Qualifiers
FEATURES
source
1..680
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 82.5%; Score 27.4; DB 2; Length 680;
Best Local Similarity 59.5%; Pred. No. 0.13;
Matches 25; Conservative 7; Mismatches 10; Indels 0; Gaps 0;
QY 1 GCNCAYCNGTGGCCNTCNCNTAYAARGAYTAYGAGTNAAR 42
|||:||||| |||||:|||||:|||||: |||
Db 633 GCCCAGCTTGGCCCAACCTTACAAAGACTATGAANGTAA 674
|||:||||| |||||:|||||:|||||: |||
RESULT 24
AC149246 224653 bp DNA linear HTG 19-JUN-2004
LOCUS
DEFINITION Orolemur garnettii clone CH256-3707, WORKING DRAFT SEQUENCE, 8
ordered pieces.
ACCESSION AC149246
VERSION AC149246.2 GI:48958688
KEYWORDS HTG; HTGS PHASE2; HTGS DRAFT.
SOURCE Orolemur garnettii (small-eared galago)
ORGANISM
Orolemur garnettii
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Strepsirrhini;
Galagonidae; Orolemur.
REFERENCE
1 (bases 1 to 224653)
AUTHORS Antonellis,A., Ayele,K., Benjamin,B., Blakesley,R.W.,
```

```
Bouffard,G.G., Brinkley,C., Brooks,S., Chu,G., Coleman,B.,
Coleman,H., Daki,N., Engle,J., Guan,X., Gupta,J., Haghighi,P.,
Han,J., Hansen,N., Ho,S.-L., Hu,P., Hurle,B., Idol,J.R., Jones,C.,
Karlsen,E., Kim,H., Kwong,P., Laric,P., Larson,S., Lee-Lin,S.-Q.,
Legaspi,R., Maskeri,B., Maduro,Q.L., Maduro,V.B., Margulies,E.H.,
Masiello,C., Maskeri,B., McDowell,J., Mullikin,J.C., Paquitigan,C.,
Park,M., Portnoy,M.E., Prasad,A., Puri,O., Reddix-Dugue,N.,
Schandler,K., Schueler,M.G., Shah,K., Sison,C., Stantirpop,S.,
Thomas,J.W., Thomas,P.J., Tsipouri,V., Vogt,J.L., Wetherby,K.D.,
Young,A. and Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 224653)
Green,E.D.
Direct Submission
Submitted (27-MAY-2004) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA
3 (bases 1 to 224653)
Green,E.D.
Direct Submission
Submitted (19-JUN-2004) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA
On Jun 19, 2004 this sequence version replaced gi:47716538.
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_zoo@hgri.nih.gov
----- Project Information
Center project name: gpg
Center clone name: 037007
The sequence data in this record represents an 'enhanced'
version of a Phase 2 submission. Specifically, the indicated
order and orientation of each sequence contig has been
established using one or more of the following: read-pair
data from individual subclones, overlaps with neighboring
clones, alignment with available reference sequence (e.g.,
human), and/or confirmation by PCR testing. In addition,
the sequence assembly is based on at least 8X average
coverage in Q20 bases and has been reviewed to rule out
gross misassemblies, the low-quality ends of sequence
contigs have been trimmed away, and each base is associated
with a Phrap-derived quality score.
----- Summary Statistics
Sequencing vector: plasmid, n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 223254 bases at least Q40
Consensus quality: 223828 bases at least Q30
Consensus quality: 223915 bases at least Q20
Insert size: 233000; agarose-fp
Insert size: 223953; sum-of-contigs
Quality coverage: 11.53x in Q20 bases; agarose-fp
Quality coverage: 12.00x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 8440: contig of 8440 bp in length
* 8441: gap of unknown length
* 21103: contig of 12563 bp in length
* 8541
* 21104: gap of unknown length
* 21203: contig of 15125 bp in length
* 36328: gap of unknown length
* 36329
* 36428: gap of unknown length
* 36429
* 45563: contig of 9135 bp in length
* 45564
* 45663: gap of unknown length
```



Assembly program: Phrap: version 990315  
 Consensus quality: 200171 bases at least Q40  
 Consensus quality: 200668 bases at least Q30  
 Consensus quality: 200809 bases at least Q20  
 Insert size: 201600; sum-of-contigs  
 Quality coverage: 6.5x in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 6 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1 1070: contig of 1070 bp in length  
 \* 1170: gap of unknown length  
 \* 1171 12362: contig of 11192 bp in length  
 \* 12363 12462: gap of unknown length  
 \* 12463 42432: contig of 29970 bp in length  
 \* 42433 42532: gap of unknown length  
 \* 42533 76804: contig of 34272 bp in length  
 \* 76805 76904: gap of unknown length  
 \* 76905 140534: contig of 63630 bp in length  
 \* 140535 140635: gap of unknown length  
 \* 140635 202050: contig of 61416 bp in length.

#### FEATURES

source  
 1..202050  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="10"  
 /clone="RP11-141A16"  
 /clone\_lib="RPCI-11"  
 1..1070  
 /note="assembly\_name:Contig8"  
 1071..1170  
 /estimated\_length=unknown  
 1171..12362  
 /note="assembly\_name:Contig20"  
 12363..12462  
 /estimated\_length=unknown  
 12463..42432  
 /note="assembly\_name:Contig21"  
 clone\_end:T7"  
 42433..42532  
 /estimated\_length=unknown  
 42533..76804  
 /note="assembly\_name:Contig22"  
 76805..76904  
 /estimated\_length=unknown  
 76905..140534  
 /note="assembly\_name:Contig23"  
 140535..140634  
 /estimated\_length=unknown  
 140635..202050  
 /note="assembly\_name:Contig24"

#### ORIGIN

Query Match 66.3%; Score 22; DB 12; Length 202050;  
 Best Local Similarity 52.4%; Pred. No. 1.7e+02;  
 Matches 22; Conservative 7; Mismatches 13; Indels 0; Gaps 0;

QY 1 GCNCAVCGTGGCNCNTCNCTAYAAAGATYAGTGTTAA 42  
 |||||  
 Db 155822 GCCATTTCGGACATCTCCACAAAGTCTACGGACTTTAA 155863

#### RESULT 27

AC018977/c  
 LOCUS  
 DEFINITION Homo sapiens chromosome 10 clone RP11-386C23, complete sequence.  
 ACCESSION AC018977

#### VERSION KEYWORDS SOURCE ORGANISM

AC018977.9 GI:23499518

HTG.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homiidae; Homo.

1 (bases 1 to 222810)

Smith, D.R.

Genome Therapeutics Corporation Sequencing Center: Human Genome

Sequence Data

Unpublished

2 (bases 1 to 222810)

Smith, D.R.

Direct Submission

Submitted (25-DEC-1999) Genome Therapeutics Corporation, 100 Beaver

Street, Waltham, MA 02453, USA

3 (bases 1 to 222810)

Smith, D.R.

Direct Submission

Submitted (21-AUG-2001) Genome Therapeutics Corporation, 100 Beaver

Street, Waltham, MA 02453, USA

4 (bases 1 to 222810)

Smith, D.R.

Direct Submission

Submitted (27-MAR-2002) Genome Therapeutics Corporation, 100 Beaver

Street, Waltham, MA 02453, USA

5 (bases 1 to 222810)

Smith, D.R.

Direct Submission

Submitted (28-MAR-2002) Genome Therapeutics Corporation, 100 Beaver

Street, Waltham, MA 02453, USA

6 (bases 1 to 222810)

Smith, D.R.

Direct Submission

Submitted (09-APR-2002) Genome Therapeutics Corporation, 100 Beaver

Street, Waltham, MA 02453, USA

7 (bases 1 to 222810)

Smith, D.R.

Direct Submission

Submitted (04-OCT-2002) Genome Therapeutics Corporation, 100 Beaver

Street, Waltham, MA 02453, USA

On Oct 4, 2002 this sequence version replaced gi:19744961.

Location/Qualifiers

source

1..222810  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="10"  
 /clone="RP11-386C23"  
 /clone\_lib="RPCI-11"

#### ORIGIN

Query Match 66.3%; Score 22; DB 5; Length 222810;  
 Best Local Similarity 52.4%; Pred. No. 1.7e+02;  
 Matches 22; Conservative 7; Mismatches 13; Indels 0; Gaps 0;

QY 1 GCNCAVCGTGGCNCNTCNCTAYAAAGATYAGTGTTAA 42  
 |||||  
 Db 14166 GCCATTTCGGACATCTCCACAAAGTCTACGGACTTTAA 14125

#### RESULT 28

AC093902

LOCUS

DEFINITION Homo sapiens BAC clone RP11-732B24 from 4, complete sequence.

ACCESSION AC093902 AC026505

VERSION AC093902.3 GI:16077064

KEYWORDS

SOURCE

ORGANISM

Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;



```
repeat_region 12977..13263
/rpt_family="L2"
repeat_region 13510..13532
/rpt_family="AT_rich"
repeat_region 13603..13691
/rpt_family="L1"
repeat_region 13895..13920
/rpt_family="AT_rich"
repeat_region 14918..14951
/rpt_family="(T)n"
repeat_region 14922..15232
/rpt_family="Alu"
repeat_region 15697..15719
/rpt_family="AT_rich"
repeat_region 18777..18841
/rpt_family="MIR"
repeat_region 18881..18913
/rpt_family="(CATATA)n"
repeat_region 19141..19163
/rpt_family="AT_rich"
repeat_region 19255..19285
/rpt_family="AT_rich"
repeat_region 19381..19436
/rpt_family="(TA)n"
repeat_region 19771..19794
/rpt_family="AT_rich"
repeat_region 19974..20278
/rpt_family="Alu"
repeat_region 20252..20286
/rpt_family="(CAAAA)n"
repeat_region 20413..20478
/rpt_family="AT_rich"
repeat_region 20578..20697
/rpt_family="L1"
repeat_region 20698..20913
/rpt_family="Alu"
repeat_region 20892..20926
/rpt_family="(TAA)n"
repeat_region 20914..21078
/rpt_family="L1"
repeat_region 21110..21314
/rpt_family="L1"
repeat_region 21121..21218
/rpt_family="A-rich"
repeat_region 21273..21295

Query Match 65.1%; Score 21.6; DB 5; Length 179591;
Best Local Similarity 56.8%; Pred. No. 2.6e+02;
Matches 21; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY 6 YGNTGGCCNTCCNTAYARGAYTAYGARGTNAAR 42
Db 108898 TGCTGGCTGCTCCCTACGAGGAGTGAAGTAAAG 108934

RESULT 29
AL606745/c
LOCUS Mouse DNA sequence from clone RP23-32L6 on chromosome 3, complete
DEFINITION sequence.
ACCESSION AL606745
VERSION AL606745.11 GI:24527386
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
JOURNAL Sciurognathi; Muroidea; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 192332)
REFERENCE Direct Submission
AUTHORS Corby,N.
JOURNAL Submitted (01-NOV-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

repeat_region 12977..13263
/rpt_family="L2"
repeat_region 13510..13532
/rpt_family="AT_rich"
repeat_region 13603..13691
/rpt_family="L1"
repeat_region 13895..13920
/rpt_family="AT_rich"
repeat_region 14918..14951
/rpt_family="(T)n"
repeat_region 14922..15232
/rpt_family="Alu"
repeat_region 15697..15719
/rpt_family="AT_rich"
repeat_region 18777..18841
/rpt_family="MIR"
repeat_region 18881..18913
/rpt_family="(CATATA)n"
repeat_region 19141..19163
/rpt_family="AT_rich"
repeat_region 19255..19285
/rpt_family="AT_rich"
repeat_region 19381..19436
/rpt_family="(TA)n"
repeat_region 19771..19794
/rpt_family="AT_rich"
repeat_region 19974..20278
/rpt_family="Alu"
repeat_region 20252..20286
/rpt_family="(CAAAA)n"
repeat_region 20413..20478
/rpt_family="AT_rich"
repeat_region 20578..20697
/rpt_family="L1"
repeat_region 20698..20913
/rpt_family="Alu"
repeat_region 20892..20926
/rpt_family="(TAA)n"
repeat_region 20914..21078
/rpt_family="L1"
repeat_region 21110..21314
/rpt_family="L1"
repeat_region 21121..21218
/rpt_family="A-rich"
repeat_region 21273..21295

Query Match 65.1%; Score 21.6; DB 6; Length 192332;
Best Local Similarity 56.8%; Pred. No. 2.6e+02;
Matches 21; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY 2 CMCAYGCTGGCCNTCCNTAYARGAYTAYGARGT 38
Db 172227 CACACATGGCGCTACCCACACAGGCTAATGAAGT 172191

RESULT 30
AC021062/c
LOCUS Mus musculus strain C57BL6/J chromosome 3 clone RP23-102D8, WORKING
DEFINITION DRAFT SEQUENCE, 22 unordered pieces.
ACCESSION AC021062
VERSION AC021062.16 GI:21358691
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
JOURNAL Sciurognathi; Muroidea; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 203317)
REFERENCE Direct Submission
AUTHORS Grills,G., Han,J., Montgomery,K.T., Lee,E., Long,J., Pomerantz,R.,
JOURNAL Toshihikes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A., Gordon,M.
AUTHORS and Kucherlapati,R.
TITLE 2 (bases 1 to 203317)
REFERENCE Direct Submission
AUTHORS Grills,G., Han,J., Montgomery,K.T., Lee,E., Long,J., Pomerantz,R.,
JOURNAL Toshihikes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A., Gordon,M.
AUTHORS and Kucherlapati,R.
TITLE Direct Submission
```

humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
On Nov 3, 2002 this sequence version replaced gi:24394907.  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: humquery@sanger.ac.uk  
-----

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C\_elegans/wormpep RP23-32L6 is from the RPCI-23 Mouse PAC Library constructed by the group of Pieter de Jong.

For further details see http://www.chori.org/bacpac/home.htm  
VECTOR: pBACe3.6.

FEATURES  
source  
1..192332  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="3"  
/clone="RP23-32L6"  
/clone\_lib="RPCI-23"

## ORIGIN

Query Match 65.1%; Score 21.6; DB 6; Length 192332;  
Best Local Similarity 56.8%; Pred. No. 2.6e+02;  
Matches 21; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY 2 CMCAYGCTGGCCNTCCNTAYARGAYTAYGARGT 38  
Db 172227 CACACATGGCGCTACCCACACAGGCTAATGAAGT 172191

RESULT 30  
AC021062/c  
LOCUS Mus musculus strain C57BL6/J chromosome 3 clone RP23-102D8, WORKING  
DEFINITION DRAFT SEQUENCE, 22 unordered pieces.

ACCESSION AC021062  
VERSION AC021062.16 GI:21358691  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
JOURNAL Sciurognathi; Muroidea; Muridae; Murinae; Mus.  
TITLE 1 (bases 1 to 203317)

REFERENCE Direct Submission  
AUTHORS Corby,N.  
JOURNAL Submitted (01-NOV-2002) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

## JOURNAL

Submitted (14-JAN-2000) Department of Molecular Genetics, Albert Einstein College of Medicine Genome Center, 1300 Morris Park Ave., Bronx, NY 10461, USA  
3 (bases 1 to 203317)

REFERENCE  
AUTHORS

Grills,G., Han,J., Montgomery,K.T., Lee,E., Long,J., Pomerantz,R., Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A., Gordon,M. and Kucherlapati,R.

TITLE  
JOURNAL

Direct Submission  
Submitted (08-JUN-2002) Harvard Partners Center for Genetics and Genomics, Harvard Medical School, 65 Landsdowne St, Cambridge, MA 02139, USA

## COMMENT

On Jun 8, 2002 this sequence version replaced gi:21306613.

-----Genome Center

Center: Harvard Partners Genome Center

Center Code: HPGC

Web site: <http://www.hpcgg.org/Sequence/mouse.html>

Contact: hpgc@medel.mgh.harvard.edu

-----Summary Statistics

Center project name: AAY

Sequencing vector: pUC18; L08752

Chemistry: Dye-terminator Big Dye; 100%

\*Consensus quality: 194428 at least Q20

\*Consensus quality: 191055 at least Q30

\*Consensus quality: 185439 at least Q40

Estimated insert size: agarose-FP - N/A

\*\*Estimated insert size: 202897 - sum-of-contigs

Quality coverage: agarose-FP - N/A

Quality coverage: 4.3 x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: This is a 'working draft' sequence. It currently consists of 22 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. \* This record will be updated with the finished sequence. \* as soon as it is available and the accession number will be preserved.

1 24854: contig of 24854 bp in length  
\* 24855 24874: gap of unknown length  
\* 24875 45360: contig of 20486 bp in length  
\* 45361 45380: gap of unknown length  
\* 45381 62264: contig of 16884 bp in length  
\* 62265 62284: gap of unknown length  
\* 62285 81433: contig of 19149 bp in length  
\* 81434 97133: contig of 15680 bp in length  
\* 97134 97153: gap of unknown length  
\* 97154 111163: contig of 14010 bp in length  
\* 111164 111183: gap of unknown length  
\* 11184 125606: contig of 14422 bp in length  
\* 125606 137365: gap of unknown length  
\* 137365 137384: gap of unknown length  
\* 137385 149894: contig of 12510 bp in length  
\* 149895 149915: gap of unknown length  
\* 149915 158741: contig of 8826 bp in length  
\* 158741 158761: gap of unknown length  
\* 158761 166785: contig of 8025 bp in length  
\* 166786 166806: gap of unknown length  
\* 166806 172697: contig of 5892 bp in length  
\* 172698 172717: gap of unknown length  
\* 172718 176174: contig of 3457 bp in length  
\* 176175 176194: gap of unknown length  
\* 176195 181319: contig of 5124 bp in length  
\* 181320 181339: gap of unknown length  
\* 181340 185745: contig of 4407 bp in length  
\* 185746 185766: gap of unknown length  
\* 185767 191167: contig of 5401 bp in length  
\* 191168 195048: gap of unknown length  
\* 195049 195068: contig of 3861 bp in length  
\* 195069 197304: gap of unknown length  
\* 197305 197324: contig of 2237 bp in length  
\* 197325 198964: contig of 1640 bp in length  
\* 198965 200466: contig of 1482 bp in length  
\* 200467 200487: gap of unknown length  
\* 200488 201931: contig of 1445 bp in length  
\* 201932 201951: gap of unknown length  
\* 201952 203317: contig of 1366 bp in length.

## FEATURES

## source

1..203317  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/strain="C57BL6/J"  
/db\_xref="taxon:10090"  
/chromosome="3"  
/clone="RP23-102D8"  
/sex="male"  
1..24854  
/note="assembly\_name:Contig51"  
24855..24874  
/estimated\_length=unknown  
24875..45360  
/note="assembly\_name:Contig50"  
45361..45380  
/estimated\_length=unknown  
45381..62264  
/note="assembly\_name:Contig49"  
62265..62284  
/estimated\_length=unknown  
62285..81433  
/note="assembly\_name:Contig48"  
81434..81453  
/estimated\_length=unknown  
81454..97133  
/note="assembly\_name:Contig47"  
97134..97153  
/estimated\_length=unknown  
97154..111163  
/note="assembly\_name:Contig46"  
111164..111183  
/estimated\_length=unknown  
111184..125605  
/note="assembly\_name:Contig45"  
125606..125625  
/estimated\_length=unknown  
125626..137364  
/note="assembly\_name:Contig44"  
137365..137384  
/estimated\_length=unknown  
137385..149894  
/note="assembly\_name:Contig43  
clone\_end:T7  
vector\_side:right"  
149895..149914  
/estimated\_length=unknown  
149915..158740  
/note="assembly\_name:Contig42"  
158741..158760  
/estimated\_length=unknown  
158761..166785  
/note="assembly\_name:Contig41  
clone\_end:SP6  
vector\_side:left"  
166786..166805  
/estimated\_length=unknown  
166806..172697  
/note="assembly\_name:Contig40"  
172698..172717  
/estimated\_length=unknown  
172718..176174  
/note="assembly\_name:Contig39"  
176175..176194  
/estimated\_length=unknown  
176195..181318  
/estimated\_length=unknown





```
QY      2  CNCAYCGTCGCCNTCCNCTAYAAARGAYTAYGARTNAAR 42
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      544  CGCTCGCCTCGCGTCGCCATCTAAATATTACAGAGTTTAG 584

RESULT 32
CEVY10G11R      21062 bp      DNA      linear      INV 19-JAN-2006
LOCUS      Caenorhabditis elegans Plasmid VY10G11R, complete sequence.
DEFINITION      AL713392
ACCESSION      AL713392.1 GI:19848196
VERSION      HTG.
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
CONSTRM
TITLE
JOURNAL
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

C. elegans Sequencing Consortium
Genome sequence of the nematode C. elegans: a platform for
investigating biology
Science 282 (5396), 2012-2018 (1998)
9851916
2 (bases 1 to 21062)
Sulston,J.E.
Direct Submission
Submitted (19-JAN-2006) Nematode Sequencing Project, Sanger
Institute, Hinxton, Cambridge CB10 1SA, England and Department of
Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:
worm@sanger.ac.uk
Coding sequences below are predicted from computer analysis, using
predictions from GeneFinder (P. Green, U. Washington), and other
available information.
Current sequence finishing criteria for the C. elegans genome
sequencing consortium are that all bases are either sequenced
unambiguously on both strands, or on a single strand with both a
dye primer and dye terminator reaction, from distinct subclones.
Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is NOT necessarily the entire insert of
the specified clone. It may be shorter because we only sequence
overlapping sections once, or longer because we arrange for a small
overlap between neighbouring submissions.
For a graphical representation of this sequence and its analysis
see: - http://www.wormbase.org/perl/ace/elegans/seq/sequence?
names=VY10G11R;class=Sequence
IMPORTANT: This sequence is not the entire insert of clone
VY10G11R. It may be shorter because we only sequence overlapping
sections once, or longer because we arrange for a small overlap
between neighbouring submissions.
The start of this sequence (1..92) overlaps with the end of
sequence At161711.
The end of this sequence (20948..21062) overlaps with the start of
sequence At132952.
FEATURES
source
1..21062
Location/Qualifiers
/organism="Caenorhabditis elegans"
/mol_type="genomic DNA"
/strain="Bristol N2"
/db_xref="taxon:6239"
/clone="VY10G11R"
complement(join(18292..18332,18887..19592,19898..19993))
/locus_tag="VY10G11R.1"
complement(join(18292..18332,18887..19592,19898..19993))
/locus_tag="VY10G11R.1"
/standard_name="VY10G11R.1"
/notes="contains similarity to Pfam domain PF00069
(Eukaryotic protein kinase domain)"
/codon_start=1
/product="Hypothetical protein VY10G11R.1"
/protein_id="CAD29326.1"
/db_xref="GI:19848199"
/db_xref="GOA:Q8T3H2"
/db_xref="InterPro:IPR000719"

Query Match      64.5%; Score 21.4; DB 13; Length 21062;
Best Local Similarity 53.7%; Pred. No. 2.2e+02;
Matches 22; Conservative 6; Mismatches 13; Indels 0; Gaps 0;

QY      2  CNCAYCGTCGCCNTCCNCTAYAAARGAYTAYGARTNAAR 42
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      17651  CACACATTTTGGCCACCCCTGTATATAAGATTACAATGGCAAA 17691

RESULT 33
AP008217_262
WPCOMMENT
Sequence split into 284 fragments LOCUS AP008217 Accession AP008217
Fragment Name      Begin      End
AP008217_000      1      110000
AP008217_001      100001  210000
AP008217_002      200001  310000
AP008217_003      300001  410000
AP008217_004      400001  510000
AP008217_005      500001  610000
AP008217_006      600001  710000
AP008217_007      700001  810000
AP008217_008      800001  910000
AP008217_009      900001  1010000
AP008217_010     1000001  1110000
AP008217_011     1100001  1210000
AP008217_012     1200001  1310000
AP008217_013     1300001  1410000
AP008217_014     1400001  1510000
AP008217_015     1500001  1610000
AP008217_016     1600001  1710000
AP008217_017     1700001  1810000
AP008217_018     1800001  1910000
AP008217_019     1900001  2010000
AP008217_020     2000001  2110000
AP008217_021     2100001  2210000
AP008217_022     2200001  2310000
AP008217_023     2300001  2410000
AP008217_024     2400001  2510000
AP008217_025     2500001  2610000
AP008217_026     2600001  2710000
AP008217_027     2700001  2810000
AP008217_028     2800001  2910000
AP008217_029     2900001  3010000
AP008217_030     3000001  3110000
AP008217_031     3100001  3210000
AP008217_032     3200001  3310000
AP008217_033     3300001  3410000
AP008217_034     3400001  3510000
AP008217_035     3500001  3610000
AP008217_036     3600001  3710000
AP008217_037     3700001  3810000
AP008217_038     3800001  3910000
AP008217_039     3900001  4010000
AP008217_040     4000001  4110000
AP008217_041     4100001  4210000
AP008217_042     4200001  4310000
AP008217_043     4300001  4410000
AP008217_044     4400001  4510000
AP008217_045     4500001  4610000
AP008217_046     4600001  4710000
AP008217_047     4700001  4810000
AP008217_048     4800001  4910000
```

AP008217_049	4900001	5010000	AP008217_122	12200001	12310000
AP008217_050	5000001	5110000	AP008217_123	12300001	12410000
AP008217_051	5100001	5210000	AP008217_124	12400001	12510000
AP008217_052	5200001	5310000	AP008217_125	12500001	12610000
AP008217_053	5300001	5410000	AP008217_126	12600001	12710000
AP008217_054	5400001	5510000	AP008217_127	12700001	12810000
AP008217_055	5500001	5610000	AP008217_128	12800001	12910000
AP008217_056	5600001	5710000	AP008217_129	12900001	13010000
AP008217_057	5700001	5810000	AP008217_130	13000001	13110000
AP008217_058	5800001	5910000	AP008217_131	13100001	13210000
AP008217_059	5900001	6010000	AP008217_132	13200001	13310000
AP008217_060	6000001	6110000	AP008217_133	13300001	13410000
AP008217_061	6100001	6210000	AP008217_134	13400001	13510000
AP008217_062	6200001	6310000	AP008217_135	13500001	13610000
AP008217_063	6300001	6410000	AP008217_136	13600001	13710000
AP008217_064	6400001	6510000	AP008217_137	13700001	13810000
AP008217_065	6500001	6610000	AP008217_138	13800001	13910000
AP008217_066	6600001	6710000	AP008217_139	13900001	14010000
AP008217_067	6700001	6810000	AP008217_140	14000001	14110000
AP008217_068	6800001	6910000	AP008217_141	14100001	14210000
AP008217_069	6900001	7010000	AP008217_142	14200001	14310000
AP008217_070	7000001	7110000	AP008217_143	14300001	14410000
AP008217_071	7100001	7210000	AP008217_144	14400001	14510000
AP008217_072	7200001	7310000	AP008217_145	14500001	14610000
AP008217_073	7300001	7410000	AP008217_146	14600001	14710000
AP008217_074	7400001	7510000	AP008217_147	14700001	14810000
AP008217_075	7500001	7610000	AP008217_148	14800001	14910000
AP008217_076	7600001	7710000	AP008217_149	14900001	15010000
AP008217_077	7700001	7810000	AP008217_150	15000001	15110000
AP008217_078	7800001	7910000	AP008217_151	15100001	15210000
AP008217_079	7900001	8010000	AP008217_152	15200001	15310000
AP008217_080	8000001	8110000	AP008217_153	15300001	15410000
AP008217_081	8100001	8210000	AP008217_154	15400001	15510000
AP008217_082	8200001	8310000	AP008217_155	15500001	15610000
AP008217_083	8300001	8410000	AP008217_156	15600001	15710000
AP008217_084	8400001	8510000	AP008217_157	15700001	15810000
AP008217_085	8500001	8610000	AP008217_158	15800001	15910000
AP008217_086	8600001	8710000	AP008217_159	15900001	16010000
AP008217_087	8700001	8810000	AP008217_160	16000001	16110000
AP008217_088	8800001	8910000	AP008217_161	16100001	16210000
AP008217_089	8900001	9010000	AP008217_162	16200001	16310000
AP008217_090	9000001	9110000	AP008217_163	16300001	16410000
AP008217_091	9100001	9210000	AP008217_164	16400001	16510000
AP008217_092	9200001	9310000	AP008217_165	16500001	16610000
AP008217_093	9300001	9410000	AP008217_166	16600001	16710000
AP008217_094	9400001	9510000	AP008217_167	16700001	16810000
AP008217_095	9500001	9610000	AP008217_168	16800001	16910000
AP008217_096	9600001	9710000	AP008217_169	16900001	17010000
AP008217_097	9700001	9810000	AP008217_170	17000001	17110000
AP008217_098	9800001	9910000	AP008217_171	17100001	17210000
AP008217_099	9900001	10010000	AP008217_172	17200001	17310000
AP008217_100	10000001	10110000	AP008217_173	17300001	17410000
AP008217_101	10100001	10210000	AP008217_174	17400001	17510000
AP008217_102	10200001	10310000	AP008217_175	17500001	17610000
AP008217_103	10300001	10410000	AP008217_176	17600001	17710000
AP008217_104	10400001	10510000	AP008217_177	17700001	17810000
AP008217_105	10500001	10610000	AP008217_178	17800001	17910000
AP008217_106	10600001	10710000	AP008217_179	17900001	18010000
AP008217_107	10700001	10810000	AP008217_180	18000001	18110000
AP008217_108	10800001	10910000	AP008217_181	18100001	18210000
AP008217_109	10900001	11010000	AP008217_182	18200001	18310000
AP008217_110	11000001	11110000	AP008217_183	18300001	18410000
AP008217_111	11100001	11210000	AP008217_184	18400001	18510000
AP008217_112	11200001	11310000	AP008217_185	18500001	18610000
AP008217_113	11300001	11410000	AP008217_186	18600001	18710000
AP008217_114	11400001	11510000	AP008217_187	18700001	18810000
AP008217_115	11500001	11610000	AP008217_188	18800001	18910000
AP008217_116	11600001	11710000	AP008217_189	18900001	19010000
AP008217_117	11700001	11810000	AP008217_190	19000001	19110000
AP008217_118	11800001	11910000	AP008217_191	19100001	19210000
AP008217_119	11900001	12010000	AP008217_192	19200001	19310000
AP008217_120	12000001	12110000	AP008217_193	19300001	19410000
AP008217_121	12100001	12210000	AP008217_194	19400001	19510000



except on the rare occasion of the clone being a YAC.  
RP11-54H19 is from the library RPCI-11.1 constructed by the group  
of Pieter de Jong. For further details see  
<http://www.chori.org/bacpac/home.htm>  
VECTOR: pBACe3.6.

## FEATURES

source	Location/Qualifiers	gene
	1. .164168 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="1" /clone="RP11-54H19" /clone_lib="RPCI-11.1"	gene
misc_feature	1	
gene	/note="Clone left end: RP11-54H19" /locus tag="RP11-54H19.1-012" /gene="LMNA"	mRNA
mRNA	join(AL355388.30:176665..176864,2358..2813) /locus tag="RP11-54H19.1-012" /gene="LMNA"	gene
gene	/note="match: ESTs: Em:AI929169.1" join(2395..2962,18305..18461,22091..22216,22493..22663, 22875..23000,23589..23809,23902..24124,24609..24716, 24801..24920,25342..25431,26176..26445,26768..27777) /gene="LMNA"	gene
mRNA	/locus tag="RP11-54H19.1-001" join(2395..2962,18305..18461,22091..22216,22493..22663, 22875..23000,23589..23809,23902..24124,24609..24716, 24801..24920,25342..25431,26176..26445,26768..27777) /gene="LMNA"	gene
gene	/locus tag="RP11-54H19.1-001" /product="lamin A/C" /note="match: ESTs: Em:AI929169.1" join(2395..2962,18305..18461,22091..22216,22493..22663, 22875..23000,23589..23809,23902..24124,24609..24716, 24801..24920,25342..25431,26176..26445,26768..27777) /gene="LMNA"	gene
mRNA	/locus tag="RP11-54H19.1-001" /product="lamin A/C" /note="match: ESTs: Em:AI872233.1 Em:BF035195.1 Em:BG420208.1 Em:BG701671.1 Em:BM468449.1 Em:BO940951.1 match: CDNAS: Em:AK004619.2 Em:AK026584.1 Em:BC014507.1 Em:BC018863.2 Em:M13452.1 Em:X03444.1" join(AL355388.30:176258..176343, AL355388.30:176628..176864,AL355388.30:184151..184262, 2395..2962,18305..18461,22091..22216,22493..22663, 22875..23000,23589..23809,23902..24124,24609..24716, 24801..24920,25342..25553) /gene="LMNA"	mRNA
gene	/locus tag="RP11-54H19.1-002" /product="lamin A/C" /note="match: CDNAS: Em:AK056143.1 Em:X03445.1" join(2399..2577,2651..2962,18305..18461,22091..22216, 22493..22663,22875..23000,23589..23809,23902..24124, 24609..24716,24801..24920,25342..25431,26176..26445, 26768..26991) /gene="LMNA"	gene
mRNA	/locus tag="RP11-54H19.1-002" /product="lamin A/C" /note="match: CDNAS: Em:AK056143.1 Em:X03445.1" join(2399..2577,2651..2962,18305..18461,22091..22216, 22493..22663,22875..23000,23589..23809,23902..24124, 24609..24716,24801..24920,25342..25431,26176..26445, 26768..26991) /gene="LMNA"	mRNA
gene	/locus tag="RP11-54H19.1-008" /product="lamin A/C" /note="match: CDNAS: Em:AK122732.1" join(AL355388.30:176602..176864,2401..2946) /gene="LMNA"	gene
mRNA	/locus tag="RP11-54H19.1-009" /product="lamin A/C" /note="match: CDNAS: Em:AK122732.1" join(AL355388.30:176602..176864,2401..2946) /gene="LMNA"	mRNA

/locus tag="RP11-54H19.1-009"  
/product="lamin A/C"  
/note="match: ESTs: Em:BQ431561.1"  
join(AL355388.30:176295..176343,  
AL355388.30:176628..176864,2401..2819)  
/gene="LMNA"  
/locus tag="RP11-54H19.1-011"  
join(AL355388.30:176295..176343,  
AL355388.30:176628..176864,2401..2819)  
/gene="LMNA"  
/locus tag="RP11-54H19.1-011"  
/product="lamin A/C"  
/note="match: ESTs: Em:AUI30227.1"  
join(2410..2962,18305..18461,22091..22216,22493..22663,  
22875..23000,23589..23809,23902..24124,24609..24716,  
24801..24920,25342..25431,26176..26295,26768..27005)  
/gene="LMNA"  
/locus tag="RP11-54H19.1-004"  
join(2410..2962,18305..18461,22091..22216,22493..22663,  
22875..23000,23589..23809,23902..24124,24609..24716,  
24801..24920,25342..25431,26176..26295,26768..27005)  
/gene="LMNA"  
/locus tag="RP11-54H19.1-004"  
/product="lamin A/C"  
/note="match: CDNAS: Em:AY357727.1"  
join(2414..2962,18305..18461,22091..22216,22493..22663,  
22875..23000,23589..23809,23902..24080,25391..25554)  
/gene="LMNA"  
/locus tag="RP11-54H19.1-003"  
/product="lamin A/C"  
/note="match: CDNAS: Em:BC033088.1"  
join(2607..2962,18305..18461,22091..22216,22493..22663,  
22875..23000,23589..23809,23902..24124,24609..24716,  
24801..24920,25342..25431,26176..26445,26768..26794)  
/gene="LMNA"  
/locus tag="RP11-54H19.1-008"  
/standard\_name="OTTHUMP00000015848"  
/note="match: proteins: Sw:P02545 Sw:P48678 Sw:P48679  
Tr:Q91WF2 Tr:Q9DC21"  
/codon\_start=1  
/product="lamin A/C"  
/protein\_id="CAI15522.1"  
/db\_xref="GI:55957498"  
/db\_xref="GOA:Q5TCJ2"  
/db\_xref="InterPro:IPR001322"  
/db\_xref="UniProtKB/TrEMBL:Q5TCJ2"  
/translation="METPSQRATRSQAQASSTPLSPTRITRLQEKEDLOELNDRLAV  
YIDVRSLTENAGLRITSEEVSVSGIKAAAYEALGDKARKTLDVSAKERARL  
QELSKVRFEFKELKARNTKQEDLIAAQRKLDALEALNSKEAALSTALSEKRTLEG  
ELHDLRGVAKLAALGAQKQLQDEMLRVDADENRLQTKWEILDFOKNTYSEELRET  
KRRLTVLVDNGKQRESFRLQDALQELRAEDQVOYKKELEKTSYAKLDNAQ  
SAERNSLVGAHEELQOQSRIRIDSLSAQLSQKQLAAKEALRDLEDSIARERDTS  
RRLAKERMAEMARMQOQLDEYQELLDIKALDMEIHYRKLKLSGEERLRLSPS  
PYSQRGRASSHSSQTQGGSVTKRKLESTSESSFSQHARTSGRVAAVEEDEGR  
FVLRNKNEDQSMGNWQIKRQNGDDPLLTFRFPKFTLKAGQVTTIWAAGATHHP  
PDLRWKANTQCGNSLRATLINSIGEEVAMRKLRSVTVVEDEDEDGDLHHRH  
GHCSSSGPPAEYNLSRTVLCCTCGPADKASGSGAQVGGPISGSSASSVTVTR  
SVRSVGGSGSGFGDNLVTRSYLLGNSSPTQSPONCSIN"  
join(2607..2962,18305..18461,22091..22216,22493..22663,  
22875..23000,23589..23809,23902..24124,24609..24716,  
24801..24920,25342..25431,26176..26295,26768..26794)  
/gene="LMNA"  
/locus tag="RP11-54H19.1-004"  
/standard\_name="OTTHUMP00000015845"  
/note="match: proteins: Tr:AAR29466"  
/codon\_start=1  
/product="lamin A/C"  
/protein\_id="CAI15523.1"



```
number X04143"
/note="Human osteocalcin precursor (gamma-carboxyglutamic acid). Accession number P02818"
/misc_feature
/note="Human osteocalcin gene, 5' end, and promoter region (M34013)"
repeat_region
complement(35719..36177)
/rpt_family="Alu"
misc_feature
complement(join(36901..37900,39601..40600,41345..41500,42267..42400,45881..46000,49501..49782,52201..53200))
/note="Homo sapiens polyamine modulated factor-1 (AF141309)"
repeat_region
38224..38381
/rpt_family="L1"
repeat_region
38438..38781
/rpt_family="Alu"
repeat_region
38735..39038
/rpt_family="Alu"
repeat_region
39252..39367
/rpt_family="Alu"
repeat_region
39395..39520
/rpt_family="L1"
repeat_region
40725..41022
/rpt_family="Alu"
repeat_region
41058..41317
/rpt_family="Alu"
repeat_region
41593..41667
/rpt_family="MIR2"
repeat_region
41673..41977
/rpt_family="Alu"
repeat_region
complement(42395..42527)
/rpt_family="Alu"
repeat_region
42562..42704
/rpt_family="Alu"
misc_feature
complement(42789..42889)
/note="GRAIL 2 excellent exon, frame 0"
repeat_region
43007..43300
/rpt_family="Alu"
repeat_region
43395..43782
/rpt_family="Alu"
misc_feature
complement(44092..44197)
/note="GRAIL 2 excellent exon, frame 1"
repeat_region
complement(44770..45050)
/rpt_family="Alu"
repeat_region
complement(45070..45248)
/rpt_family="Alu"
repeat_region
complement(45264..45564)
/rpt_family="Alu"
repeat_region
complement(45567..45701)
/rpt_family="Alu"
repeat_region
46611..46910
/rpt_family="Alu"
repeat_region
47258..47576
/rpt_family="Alu"
misc_feature
complement(48074..48187)
/note="GRAIL 2 excellent exon, frame 0"
repeat_region
48208..48545
/rpt_family="Alu"
repeat_region
48512..48678
/rpt_family="Alu"
repeat_region
48669..48996
/rpt_family="Alu"
repeat_region
complement(49023..49273)
/rpt_family="Alu"
repeat_region
49717..49867
/rpt_family="MIR"
repeat_region
50215..50644
/rpt_family="Alu"
repeat_region
51016..51174
/rpt_family="L1"
repeat_region
51301..51581
/rpt_family="Alu"
```

```
repeat_region complement(51693..51987)
/rpt_family="Alu"

Query Match 64.5% Score 21.4; DB 5; Length 164179;
Best Local Similarity 57.9%; Pred. No. 3.2e+02;
Matches 22; Conservative 5; Mismatches 11; Indels 0; Gaps 0;

QY 1 GCNCAYGCGTGGCNCNTGCCNTAAGAAGTAYGARGT 38
|||
Db 127938 GCACACATCTTCCTCTCTTAAATATCATGAAGT 127901
|||

RESULT 37
AC158014 186979 bp DNA linear HTG 01-JUL-2005
LOCUS Bos taurus clone CH240-271C13, *** SEQUENCING IN PROGRESS ***, 6
DEFINITION unorderd pieces.
AC158014
ACCESSION AC158014.2 GI:68300579
VERSION HTG: HTGS PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
KEYWORDS Bos taurus (cattle)
SOURCE Bos taurus
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
Pecora; Bovidae; Bovinae; Bos.
1 (bases 1 to 186979)
Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Ayoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Cesar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hughes,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuewa,L., Loulseghe,H., Lozada,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Parks,K.,
Nwaokeme,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Shen,H.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajda,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villasaana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczky,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
```







gap 134634..134683  
 gap /estimated\_length=50  
 143351..143400

Query Match 64.5%; Score 21.4; DB 12; Length 196244;  
 Best Local Similarity 53.7%; Pred. No. 3.3e+02;  
 Matches 22; Conservative 6; Mismatches 13; Indels 0; Gaps 0;

QY 2 CNCAYGNTGCCNTCCNTAYAAAGAYTAYGARGTNAAR 42  
 |||: ||| |||: |||: |||: |||: |||: |||:  
 Db 155585 CTCATCTCTGCCACCCCATACATATATAAGAAAGTAAAA 155625

RESULT 39

AF110497  
 LOCUS Bartonella quintana filament-A precursor (fila) gene, partial cds;  
 DEFINITION carboxy-terminal protease (ctpa) gene, complete cds; and  
 invasion-associated locus A protein (iala) gene, partial cds.

ACCESSION AF110497  
 VERSION AF110497.1 GI:4140713  
 KEYWORDS  
 SOURCE  
 ORGANISM Bartonella quintana  
 Bartonella quintana  
 Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 Bacteriellaceae; Bartonella.

REFERENCE 1 (bases 1 to 1604)  
 AUTHORS Minnick,M.F.; Smitherman,L.S. and Carroll,J.  
 TITLE A tail-specific protease from Bartonella quintana  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 1604)  
 AUTHORS Minnick,M.F., Smitherman,L.S. and Carroll,J.  
 TITLE Direct Submission  
 JOURNAL Submitted (02-DEC-1998) Division of Biological Sciences, The  
 University of Montana, Health Sciences 104, Missoula, MT  
 59812-1002, USA

FEATURES  
 source  
 1..1604  
 /organism="Bartonella quintana"  
 /mol\_type="genomic DNA"  
 /strain="Houston 90-268"  
 /db\_xref="taxon:803"  
 <1..203  
 /gene="fila"  
 <1..203  
 /gene="fila"  
 /note="putative alpha-helical surface protein; similar to  
 Bartonella bacilliformis filament-A precursor"  
 /codon\_start=3  
 /transl\_table=11  
 /product="filament-A precursor"  
 /protein\_id="AAD04179.1"  
 /db\_xref="GI:4140715"  
 /translation="NVTQGFVLGSEPLGTGMGQFIANSVALDIGKTAPMLYIEFRQ  
 GKPVPFPWQSEKTKRNQDS"  
 193..1518  
 /gene="ctpa"  
 193..1518  
 /gene="ctpa"  
 /note="tail-specific protease; similar to Bartonella  
 bacilliformis carboxy-terminal protease"  
 /codon\_start=1  
 /transl\_table=11  
 /product="carboxy-terminal protease"  
 /protein\_id="AAD04178.1"  
 /db\_xref="GI:4140714"  
 /translation="MIRKVTLLVAGLLGASSMTWQSVAAANEEDTYQQLSLFGDIF  
 ERMVQYVTPDDKKLIENAINGLMTSLDPHSSVMNAEAKDMWDS TKGFGGLGIV  
 TMEKNLVKVPSPMDTPASNAKGLAGLIDKIGVQTNGQTLEAVKRMGAGPPTPI  
 LTTIRSGVDKPLSEIKVRDIIKKVAKYRVEDDIGYLRLLRFSEQTGDIQTAIKDIO  
 SKTPQDKLGVLDLRNPGLLNQAVNSVAFLNKGEIVSTRGRKNDVTRFDEKADQ  
 DINGKPLIIVLINGSASAEIVAGALQDHRRATILGTQSGFGKSVQTVIPLGNGAL  
 RLTTALYTPAGTSIQGTGTPDIVVEOPLTPYQYKDYDKLGESELKGHIKGKQESNK  
 "

gene  
 CDS

gene	GGGSAAFVDPKDDTQTLSEAYKLRLGEWAHAAPFPDPPNVLKKDVLK"					
CDS	1560..>1604 /gene="ialA" 1560..>1604 /gene="ialA" /note="similar to Bartonella bacilliformis invasion-associated protein" /codon_start=1 /transl_table=11 /product="invasion-associated locus A protein" /protein_id="ABD04180.1" /db_xref="GI:4140716" /translation="MDTDINLKIFLIEDV"					
ORIGIN						
Query Match	63.9%; Score 21.2; DB 15; Length 1604;					
Best Local Similarity	54.8%; Pred.No. 1.7e+02;					
Matches	23; Conservative 5; Mismatches 14; Indels 0; Gaps 0;					
QY	1 GCNCAYGCNTGGCCNTCNCTAYAARGAVTAYGAGTGTAAR 42					
Db	1273 GACCAACCATGTGCTGAACAATATAAGATTACGATGTTAAAG 1314					
RESULT 40						
BX897700_01						
WPCOMMENT						
Sequence split into 16 fragments	LOCUS BX897700 Accession BX897700					
Fragment Name	Begin	End				
BX897700_00	1	110000				
BX897700_01	100001	210000				
BX897700_02	200001	310000				
BX897700_03	300001	410000				
BX897700_04	400001	510000				
BX897700_05	500001	610000				
BX897700_06	600001	710000				
BX897700_07	700001	810000				
BX897700_08	800001	910000				
BX897700_09	900001	1010000				
BX897700_10	1000001	1110000				
BX897700_11	1100001	1210000				
BX897700_12	1200001	1310000				
BX897700_13	1300001	1410000				
BX897700_14	1400001	1510000				
BX897700_15	1500001	1581384				
Continuation (2 of 16) of BX897700 from base 100001	Bartonella quintana					
Query Match	63.9%; Score 21.2; DB 15; Length 110000;					
Best Local Similarity	54.8%; Pred.No. 3.7e+02;					
Matches	23; Conservative 5; Mismatches 14; Indels 0; Gaps 0;					
QY	1 GCNCAYGCNTGGCCNTCNCTAYAARGAVTAYGAGTGTAAR 42					
Db	101554 GAGCAACCATGTGCTGAACAATATAAGATTACGATGTTAAAG 101595					
RESULT 41						
BX897700_02						
WPCOMMENT						
Sequence split into 16 fragments	LOCUS BX897700 Accession BX897700					
Fragment Name	Begin	End				
BX897700_00	1	110000				
BX897700_01	100001	210000				
BX897700_02	200001	310000				
BX897700_03	300001	410000				
BX897700_04	400001	510000				
BX897700_05	500001	610000				
BX897700_06	600001	710000				
BX897700_07	700001	810000				
BX897700_08	800001	910000				
BX897700_09	900001	1010000				
BX897700_10	1000001	1110000				
BX897700_11	1100001	1210000				

```

BX897700.12 1200001 1310000
BX897700.13 1300001 1410000
BX897700.14 1400001 1510000
BX897700.15 1500001 1581384
Continuation (3 of 16) of BX897700 from base 200001 (BX897700 Bartonella quintana str. T
Query Match 63.9%; Score 21.2; DB 15; Length 110000;
Best Local Similarity 54.8%; Pred. No. 3.7e+02;
Matches 23; Conservative 5; Mismatches 14; Indels 0; Gaps 0;

Qy 1 GCNCAVCGTCGCCNTCCNTAYARGAYTAYGARGTNAAR 42
Db 1554 GAGCAACCATTCGCTGAACAATATAAAGATTACGATGTAAG 1595

RESULT 42
AC069465
LOCUS AC069465 203946 bp DNA linear HTG 06-SEP-2000
DEFINITION Mus musculus clone RP23-23L3, WORKING DRAFT SEQUENCE, 18 unordered
pieces.
ACCESSION AC069465
VERSION AC069465.2 GI:9972306
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 203946)
AUTHORS McCombie,W.R., Baker,J.P., Bahret,A., Bal,H., Dedhia,N.N., de la
Bastide,M., Huang,E.N., King,L., Kirchoff,K.A., Miller,B.,
Nascimento,L.U., O'Shaughnessy,A.L., Preston,R.R., Rodriguez,M.A.,
Shah,R.S., Shekher,M., Spiegel,L.A., Toth,K. and Vil,M.D.
Mouse Genomic Sequence
Unpublished
REFERENCE 2 (bases 1 to 203946)
AUTHORS McCombie,W.R.
TITLE Direct Submission
JOURNAL Submitted (01-JUN-2000) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
Harbor, NY 11724, USA
COMMENT On Sep 6, 2000 this sequence version replaced gi:8134853.
----- Genome Center
Center: Lita Annenberg Hazen Genome Center, Cold Spring Harbor
Laboratory
Center code: CSHL
Web site: http://www.cshl.org/genseq
Contact: mcombie@cshl.org
----- Project Information
Center project name: RP23-23L3
Center clone name: RP23-23L3
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 30687: contig of 30687 bp in length
* 30688: gap of unknown length
* 30855: contig of 27293 bp in length
* 58148: gap of unknown length
* 58314: gap of unknown length
* 82614: contig of 24299 bp in length
* 82780: gap of unknown length
* 82781: contig of 23753 bp in length
* 106533: gap of unknown length
* 106534: contig of 16755 bp in length
* 106701: contig of 16755 bp in length
* 123456: gap of unknown length
* 123623: contig of 13904 bp in length
* 137527: gap of unknown length
* 137694: contig of 9910 bp in length

```

```

* 147604 147770: gap of unknown length
* 147771 156659: contig of 8889 bp in length
* 156660 156826: gap of unknown length
* 156827 156874: contig of 8848 bp in length
* 156875 158421: gap of unknown length
* 158422 174421: contig of 8580 bp in length
* 174422 174588: gap of unknown length
* 174589 181828: contig of 7240 bp in length
* 181829 181995: gap of unknown length
* 181996 187383: contig of 5388 bp in length
* 187384 187550: gap of unknown length
* 187551 192285: contig of 4735 bp in length
* 192286 192452: gap of unknown length
* 192453 196361: contig of 3909 bp in length
* 196362 196528: gap of unknown length
* 196529 199523: contig of 2995 bp in length
* 199524 199690: gap of unknown length
* 199691 201663: contig of 1972 bp in length
* 201663 201829: gap of unknown length
* 201830 203544: contig of 1715 bp in length
* 203545 203710: gap of unknown length
* 203711 203946: contig of 236 bp in length.
FEATURES
Location/Qualifiers
1..203946
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP23-23L3"
30688..30854
/estimated_length=unknown
58148..58314
/estimated_length=unknown
82614..82780
/estimated_length=unknown
106534..106700
/estimated_length=unknown
123456..123622
/estimated_length=unknown
137527..137693
/estimated_length=unknown
147604..147770
/estimated_length=unknown
156660..156826
/estimated_length=unknown
156875..158421
/estimated_length=unknown
174422..174588
/estimated_length=unknown
181829..181995
/estimated_length=unknown
187384..187550
/estimated_length=unknown
192286..192452
/estimated_length=unknown
196362..196528
/estimated_length=unknown
199524..199690
/estimated_length=unknown
201663..201829
/estimated_length=unknown
203545..203710
/estimated_length=unknown
ORIGIN
Query Match 63.9%; Score 21.2; DB 12; Length 203946;
Best Local Similarity 58.8%; Pred. No. 4.2e+02;
Matches 20; Conservative 6; Mismatches 8; Indels 0; Gaps 0;
Qy 4 CAYCGTGGCCNTCCNTAYARGAYTAYGARG 37
||:|||||:||||:||||:|
Db 163807 CACACCTGGCCATCACAGTACAAAACATATCAGG 163840

```









```

/chrnosome="3"
/map="3p"
/clone="RP11-350A17"

Query Match      63.3%; Score 21; DB 5; Length 152580;
Best Local Similarity 51.2%; Pred. No. 5e+02;
Matches 21; Conservative 7; Mismatches 13; Indels 0; Gaps 0;

QY 2 CMCAYCCTGGCCCTCCNCCNTAATGAAGATYAGTGTAAR 42
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 151951 CTTCTCTCAGCTTCCCATGTGAAGACGATGAGGTAA 151911

RESULT 48
AC024219
LOCUS      AC024219      154010 bp      DNA      linear      HTG 06-SEP-2000
DEFINITION Homo sapiens chromosome 3 clone RP11-350A17, WORKING DRAFT
SEQUENCE, 13 unordered pieces.
AC024219
AC024219.11 GI:9438327
VERSION    HTG; HTGS PHASE1; HTGS_DRAFT.
KEYWORDS   Homo sapiens (human)
SOURCE     Homo sapiens
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homnidae; Homo.
REFERENCE  1 (bases 1 to 154010)
AUTHORS   Muzny,D.M., Adams,C., Bailey,M., Barbara,J., Blankenburg,K.,
            Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
            Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
            David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
            Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
            Forcum-Taney,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
            Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hoques,M.,
            Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
            Kelly,S., Kondejewski,N., Kong,Y., Kovari,C., Leal,B., Li,Z.,
            Lichtarge,O., Liu,W., Logan,O., Lozado,R.J., Lu,J.,
            Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M.,
            Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S.,
            Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L.,
            Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,
            Shan,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sucgang,R.,
            Tabor,P., Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wahbah,M.,
            Watlington,S., Weinstein,G., Weinstein,I.R., Williamson,A.,
            Worley,K., Wren,J., Wrensford,G., Yu,W., Zhou,X., Nelson,D. and
            Gibbs,R.
TITLE      Direct Submission
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 154010)
AUTHORS     Worley,K.C.
TITLE       Direct Submission
JOURNAL     Submitted (28-FEB-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 25, 2000 this sequence version replaced gi:8699729.
            ----- Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            ----- Project Information
            Center project name: HAIE
            Center clone name: RP11-350A17
            ----- Summary Statistics
            Sequencing vector: M13; L08821
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990329
            Consensus quality: 140016 bases at least Q40
            Consensus quality: 145094 bases at least Q30
            Consensus quality: 147945 bases at least Q20
            Estimated insert size: 150053; sum-of-contigs estimation
            Quality coverage: 0x in Q20 bases; agarose-gel estimation

-----
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
* NOTE: Estimated insert size may differ from sequence length
  (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
  * consists of 13 contigs. The true order of the pieces
  * is not known and their order in this sequence record is
  * arbitrary. Gaps between the contigs are represented as
  * runs of N, but the exact sizes of the gaps are unknown.
  * This record will be updated with the finished sequence
  * as soon as it is available and the accession number will
  * be preserved.
* 1 28396: contig of 28396 bp in length
* 28397
* 28496: gap of unknown length
* 58151: contig of 29655 bp in length
* 58152
* 58251: gap of unknown length
* 58252
* 77456: contig of 19205 bp in length
* 77457
* 77556: gap of unknown length
* 77557
* 94589: contig of 17033 bp in length
* 94590
* 94689: gap of unknown length
* 108289: contig of 13600 bp in length
* 108290
* 108389: gap of unknown length
* 108390
* 119719: contig of 11330 bp in length
* 119720
* 119819: gap of unknown length
* 128258: contig of 8439 bp in length
* 128259
* 128358: gap of unknown length
* 128359
* 135273: contig of 6915 bp in length
* 135274
* 135374: contig of 6399 bp in length
* 141772: gap of unknown length
* 141773
* 141772: contig of 5396 bp in length
* 147268: contig of 5396 bp in length
* 147269
* 147368: gap of unknown length
* 147369
* 149978: contig of 2610 bp in length
* 149979
* 150079: gap of unknown length
* 150079
* 152110: contig of 2032 bp in length
* 152111
* 152210: gap of unknown length
* 152211
* 154010: contig of 1800 bp in length.
            Location/Qualifiers
            1. .154010
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="3"
            /clone="RP11-350A17"
            28397..28496
            /estimated_length=unknown
            58152..58251
            /estimated_length=unknown
            77457..77556
            /estimated_length=unknown
            94590..94689
            /estimated_length=unknown
            108290..108389
            /estimated_length=unknown
            119720..119819
            /estimated_length=unknown
            128259..128358
            /estimated_length=unknown
            135274..135373
            /estimated_length=unknown
            141773..141872
            /estimated_length=unknown
            147269..147368
            /estimated_length=unknown
            149979..150078
            /estimated_length=unknown
            152111..152210
            /estimated_length=unknown

ORIGIN
Query Match      63.3%; Score 21; DB 12; Length 154010;
Best Local Similarity 51.2%; Pred. No. 5e+02;
Matches 21; Conservative 7; Mismatches 13; Indels 0; Gaps 0;

```



## JOURNAL

## COMMENT

2 CNAAYGNTGCCNCCNTAYAGATYAGTGNAAR 42  
 Db 142502 CTCTCTCAGCCTTCCCATGTAAAGACGATGAGTTAA 142542

AC172585 157590 bp DNA linear HTG 24-JAN-2006  
 Bos taurus clone CH240-250J10, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 16  
 unordered pieces.

AC172585  
 HTG; HTGS PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.

Bos taurus (cattle)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;  
 Pecora; Bovidae; Bovinae; Bos.

1 (bases 1 to 157590)

REFERENCE  
 AUTHORS

Muzny, D., Adams, C., Agbai, I., Allen, C., Alabrooks, S., Archer, P.,  
 Arredondo, H., Bandaranaike, D., Bangura, L., Beltran, B., Beltran, R.,  
 Beraducci, A., Biswal, K., Blyth, P., Bonham, H., Buhay, C., Burch, P.,  
 Cadoree, I., Canada, A., Cardenas, V., Carter, K., Cavazos, I.,  
 Chacko, J., Chahrour, M., Chavez, D., Chen, A., Chen, G., Chen, R.,  
 Cheng, M.-T., Chu, J., Clerc, K., Cockrell, R., Coyle, M., Cree, A.,  
 Curry, S., Dai, W., Davila, M.L., Davis, C., Davy-Carroll, L., De  
 Anda, C., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H.,  
 Donlin, J., McCauley, S., Dugan-Rocha, S., Dunn, A., Durbin, K.,  
 Dziuda, D., Egan, A., Escotto, M., Espinosa, V., Eugene, C., Fa, M.,  
 Fernandez, S., Fernando, P., Flagg, N., Forbes, L., Foster, P.,  
 Fowler, G., Fu, Q., Fuh, E., Garcia, A., Garcia, R., Garner, T.,  
 Gaskin, C., Gench, S., Ghose, S., Gill, R., Gonzalez, D.,  
 Gonzalez-Garay, M., Guevara, W., Holder, M., Haaland, W., Haebler, K.,  
 Hall, B., Hamid, H., Hamilton, K., Harbes, B., Harris, R., Havlak, P.,  
 Hawes, A., Hawkins, E., Hayes, S., Hemphill, L., Hernandez, J.,  
 Hines, S., Hitchens, M., Hodgson, A., Hogues, M., Hollins, B.,  
 Howell, L.T., Hulyk, S., Hume, J., Imo, K., Jackson, A., Jackson, L.,  
 Jacob, L., Jiang, H., Johnson, B., Johnson, R., Kalafus, K., Kelly, S.,  
 Keys, T., Khan, Z., King, L., Kovar, C., Kowis, A., Kowis, C., Lara, F.,  
 Leal, S., Lee, K., Lee, S., Legall, F.I., Lemon, S., Lewis, L., Li, B.,  
 Li, Y., Li, Z., Linell, M., Liu, W., Liu, Y.-S., Liu, Y., Liyanage, D.,  
 London, P., Lopez, J., Lorensuhalwa, L., Lozano, R., Luk, T., Madu, R.,  
 Maheshwari, M., Mahoney, C., Malloy, K., Mansouri, D., Martinez, E.,  
 McCalland, H., McPherson, J., Mercadon, C., Metzger, M.,  
 Milosavljevic, A., Minja, E., Morgan, M., Morris, S., Munidasa, M.,  
 Murray, D., Nazareth, L., Ngo, D., Nguyen, N., Norwig-Eastaugh, E.,  
 Nott, A., Nwaokemele, O., Obregon, M., Ochi-Okorie, C., Odeh, B.,  
 Okwuonu, G., Okwuonu, K., Parker, D., Pasternak, S., Patel, B.,  
 Patel, V., Paul, H., Perez, A., Perez, L., Petrosino, J., Pham, T.,  
 Primus, E., Pu, L.-L., Puazo, M., Qin, X., Quinn, A., Quiroz, J.,  
 Rabata, D., Rachlin, E., Reigh, R., Ren, Y., Reuter, M., Richards, S.,  
 Rives, C., Rodriguez, F., Rojas, A., Ruiz, S.J., Sana, M., Sanders, W.,  
 Santibanez, J., Santos, R., Saverly, G., Scherer, S., Shen, H., Shen, Y.,  
 Sisson, I., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R.,  
 Svatek, A., Taylor, E., Taylor, T., Thomas, N., Thornton, R.,  
 Trejos, Z., Usmani, K., Vargo, C., Verduzco, D., Villasana, D., Virk, D.,  
 Volkov, A., Waldron, L., Walker, B., Wang, Q., Wang, S., Warren, J.,  
 Wei, X., Wheeler, D., Williams, G., Williams, R., Worley, K., Wright, R.,  
 Wu, J., Yakub, S., Yan, K., Yuan, Y., Yu, P., Zhang, J., Zhang, L.,  
 Zhang, Z., Zhou, J., Weinstein, G. and Gibbs, R.A.

Direct Submission

Unpublished

2 (bases 1 to 157590)

REFERENCE  
 JOURNAL

Worley, K.C.

TITLE  
 Direct Submission

Submitted (21-NOV-2005) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 157590)

REFERENCE  
 AUTHORS

CONSRM

TITLE

Bovine Genome Sequencing Consortium

Direct Submission

Submitted (24-JAN-2006) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

On Jan 24, 2006 this sequence version replaced gi:82569398.  
 The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described  
 in the feature table below represents a scaffold in the Atlas  
 assembly (a 'contig-scaffold'). Within each contig-scaffold,  
 individual sequence contigs are ordered and oriented, and separated  
 by sized gaps filled with Ns to the estimated size. The sequence  
 may extend beyond the ends of the clone and there may be sequence  
 contigs within a contig-scaffold that consist entirely of whole  
 genome shotgun sequence reads. Both end sequences and whole genome  
 shotgun sequence only contigs will be indicated in the feature  
 table.

----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: http://www.hgsc.bcm.tmc.edu/  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: FOOA  
 Center clone name: CH240-250J10  
 ----- Summary Statistics  
 Assembly program: Atlas 3.0;  
 Consensus quality: 150753 bases at least Q40  
 Consensus quality: 152702 bases at least Q30  
 Consensus quality: 154389 bases at least Q20  
 Estimated insert size: 153285; sum-of-contigs estimation  
 Quality coverage: 6x in Q20 bases; sum-of-contigs estimation  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html):  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 16 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 1894: contig of 1894 bp in length  
 \* 1895 1944: gap of 50 bp  
 \* 1945 34522: contig of 32578 bp in length  
 \* 34523 34572: gap of 50 bp in length  
 \* 34573 36767: contig of 2195 bp in length  
 \* 36768 37257: gap of 490 bp  
 \* 37258 40340: contig of 3083 bp in length  
 \* 40341 40497: gap of 157 bp  
 \* 40498 57103: contig of 16606 bp in length  
 \* 57104 57153: gap of 50 bp  
 \* 57154 82191: contig of 25038 bp in length  
 \* 82192 82241: gap of 50 bp  
 \* 82242 105056: contig of 22815 bp in length  
 \* 105057 105106: gap of 50 bp  
 \* 105107 108769: contig of 3663 bp in length  
 \* 108770 108819: gap of 50 bp  
 \* 108820 135472: contig of 26653 bp in length  
 \* 135473 135824: gap of 351 bp  
 \* 135825 149096: contig of 13273 bp in length  
 \* 149097 149196: gap of unknown length  
 \* 149197 150273: contig of 1077 bp in length  
 \* 150274 150374: gap of unknown length  
 \* 150375 152260: contig of 1887 bp in length  
 \* 152261 152361: gap of unknown length  
 \* 152362 153496: contig of 1136 bp in length  
 \* 153497 153596: gap of unknown length  
 \* 153597 154626: contig of 1030 bp in length  
 \* 154627 154727: gap of unknown length  
 \* 154728 156089: contig of 1363 bp in length  
 \* 156090 156189: gap of unknown length  
 \* 156190 157590: contig of 1401 bp in length.





## JOURNAL

Submitted (31-JUL-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Jul 31, 2001 this sequence version replaced gi:15029434.  
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email [gc-help@bcm.tmc.edu](mailto:gc-help@bcm.tmc.edu)

## COMMENT

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

## ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL:  
<http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

## QUALSTAT-REPORT.

## FEATURES

Location/Qualifiers

1..178752

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="3"

/clone="RP11-554F8"

/complement(1..10232)

/notes="overlaps bases 10232..1 of clone AC055710"

/function="clone overlap"

672..705

/rpt\_family="(TG)n"

1233..1253

/rpt\_family="AT\_rich"

1698..1894

/rpt\_family="AluSq"

2952..3265

/rpt\_family="L1PA13"

3356..3386

/rpt\_family="AT\_rich"

3562..3703

/rpt\_family="CT-rich"

4704..4734

/rpt\_family="AT\_rich"

complement(4956..5141)

/rpt\_family="MIR"

6568..6869

/standard\_name="174894"

7016..7065

/rpt\_family="A-rich"

complement(10872..10947)

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

Query Match 63.3%; Score 21; DB 5; Length 178752;

Best Local Similarity 51.2%; Pred. No. 5.1e+02;

Matches 21; Conservative 7; Mismatches 13; Indels 0; Gaps 0;

QY 2 CNYAGCCTGGCCNCCNTAYAAAGAYTAYGAGTNAAR 42

DB 97957 CTCTTCTTCTCTCTCAATATGAAGATTATGAGGACAAA 97917

## RESULT 53

AC087589/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AC087589 181736 bp DNA linear PRI 08-NOV-2002  
 Homo sapiens chromosome 3 clone RP11-1102N22 map 3p, complete  
 sequence.  
 AC087589  
 AC087589.3 GI:24270662  
 HTG.  
 Homo sapiens (human)

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homiidae; Homo.

1 (bases 1 to 181736)

Niu, Y., Bao, J., Bao, Q., Bao, W., Bian, X., Cao, T., Chen, C., Chen, J.,  
 Ding, H., Dong, W., Fan, H., Feng, X., Gong, J., Guan, Q., Gu, X., Guo, D.,  
 Guo, Z., He, L., Hu, S., Huang, F., Jin, Y., Kang, N., Li, C., Li, C.,  
 Li, F., Li, G., Li, J., Li, L., Li, S., Li, T., Liu, Y., Liu, N., Liu, B.,  
 Liu, Y., Li, W., Li, W., Li, Y., Luo, J., Luo, J., Qi, Q., Qi, X., Song, L.,  
 Song, S., Sun, M., Sun, W., Sun, Y., Tan, X., Tao, R., Wang, H., Wang, J.,  
 Wang, J., Wang, L., Wang, L., Wang, R., Wang, X., Wang, X., Wang, Y.,  
 Wang, Y., Wu, D., Wu, Q., Xie, F., Xuan, Z., Xue, Y., Yan, C., Yang, X.,  
 Yu, B., Zeng, Y., Zhang, G., Zhang, H., Zhang, H., Zhang, L., Zhang, M.,  
 Zhang, X., Zhang, X., Zhang, Y., Zhang, Y., Zhang, Z., Zhu, B., Zhu, N.,  
 Yu, J. and Yang, H.

Chromosome 3p genomic sequence

Unpublished

## TITLE

JOURNAL

REFERENCE

AUTHORS

2 (bases 1 to 181736)  
 Niu, Y., Bao, J., Bao, Q., Bao, W., Bian, X., Cao, T., Chen, C., Chen, J.,  
 Ding, H., Dong, W., Fan, H., Feng, X., Gong, J., Guan, Q., Gu, X., Guo, D.,  
 He, L., Hu, S., Huang, F., Jin, Y., Kang, N., Li, C., Li, C., Li, G.,  
 Li, J., Li, L., Li, S., Li, T., Liu, Y., Liu, N., Liu, B., Liu, Y., Li, W.,  
 Li, W., Li, Y., Luo, J., Luo, J., Qi, Q., Qi, X., Song, L., Song, S., Sun, M.,  
 Sun, W., Sun, Y., Tan, X., Tao, R., Wang, H., Wang, J., Wang, L.,  
 Wang, L., Wang, R., Wang, X., Wang, X., Wang, Y., Wu, D., Wu, Q., Xie, F.,  
 Xuan, Z., Xue, Y., Yan, C., Yang, X., Yu, B., Zeng, Y., Zhang, G.,  
 Zhang, H., Zhang, H., Zhang, L., Zhang, M., Zhang, X., Zhang, X.,  
 Zhang, Y., Zhang, Y., Zhang, Z., Zhu, B., Zhu, N., Yu, J. and Yang, H.

Direct Submission

Submitted (12-JAN-2001) Human Genomic Center, Institute of  
 Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing  
 100101, P.R.China

## REFERENCE

AUTHORS

3 (bases 1 to 181736)  
 Bao, J., Bao, Q., Bao, W., Bian, X., Cao, T., Chen, C., Chen, J., Ding, H.,  
 Dong, W., Fan, H., Feng, X., Gong, J., Guan, Q., Gu, X., Guo, D., Guo, Z.,  
 He, L., Hu, S., Huang, F., Jin, Y., Kang, N., Li, C., Li, C., Li, F.,  
 Li, G., Li, J., Li, L., Li, S., Li, T., Liu, Y., Liu, N., Liu, B., Liu, Y.,  
 Li, W., Li, W., Li, Y., Luo, J., Luo, J., Qi, Q., Qi, X., Song, L.,  
 Song, S., Sun, M., Sun, W., Sun, Y., Tan, X., Tao, R., Wang, H., Wang, J.,  
 Wang, J., Wang, L., Wang, L., Wang, R., Wang, X., Wang, X., Wang, Y.,  
 Wang, Y., Wu, D., Wu, Q., Xie, F., Xuan, Z., Xue, Y., Yan, C., Yang, X., Yu, B.,  
 Zeng, Y., Zhang, G., Zhang, H., Zhang, H., Zhang, L., Zhang, L., Zhang, M.,  
 Zhang, X., Zhang, X., Zhang, Y., Zhang, Y., Zhang, Z., Zhu, B., Zhu, N.,  
 Yu, J. and Yang, H.

Direct Submission

Submitted (01-MAY-2001) Human Genomic Center, Institute of  
 Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing  
 100101, P.R.China

## REFERENCE

AUTHORS

4 (bases 1 to 181736)  
 Niu, Y., Bao, J., Bao, Q., Bao, W., Bian, X., Cao, T., Chen, C., Chen, J.,  
 Ding, H., Dong, W., Fan, H., Feng, X., Gong, J., Guan, Q., Gu, X., Guo, D.,  
 Guo, Z., He, L., Hu, S., Huang, F., Jin, Y., Kang, N., Li, C., Li, C.,  
 Li, F., Li, G., Li, J., Li, L., Li, S., Li, T., Liu, Y., Liu, N., Liu, B.,  
 Liu, Y., Li, W., Li, W., Li, Y., Luo, J., Luo, J., Qi, Q., Qi, X., Song, L.,  
 Song, S., Sun, M., Sun, W., Sun, Y., Tan, X., Tao, R., Wang, H., Wang, J.,  
 Wang, J., Wang, L., Wang, L., Wang, R., Wang, X., Wang, X., Wang, Y.,  
 Wang, Y., Wu, D., Wu, Q., Xie, F., Xuan, Z., Xue, Y., Yan, C., Yang, X., Yu, B.,  
 Zeng, Y., Zhang, G., Zhang, H., Zhang, H., Zhang, L., Zhang, L., Zhang, M.,  
 Zhang, X., Zhang, X., Zhang, Y., Zhang, Y., Zhang, Z., Zhu, B., Zhu, N.,  
 Yu, J. and Yang, H.

Direct Submission

Submitted (23-OCT-2002) Human Genomic Center, Institute of

## REFERENCE

AUTHORS

Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing  
 100101, P.R.China  
 5 (bases 1 to 181736)  
 Niu, Y., Bao, J., Bao, Q., Bao, W., Bian, X., Cao, T., Chen, C., Chen, J.,  
 Ding, H., Dong, W., Fan, H., Feng, X., Gong, J., Guan, Q., Gu, X., Guo, D.,  
 Guo, Z., He, L., Hu, S., Huang, F., Jin, Y., Kang, N., Li, C., Li, C.,  
 Li, F., Li, G., Li, J., Li, L., Li, S., Li, T., Liu, Y., Liu, N., Liu, B.,  
 Liu, Y., Li, W., Li, W., Li, Y., Luo, J., Luo, J., Qi, Q., Qi, X., Song, L.,  
 Song, S., Sun, M., Sun, W., Sun, Y., Tan, X., Tao, R., Wang, H., Wang, J.,  
 Wang, J., Wang, L., Wang, L., Wang, R., Wang, X., Wang, X., Wang, Y.,  
 Wang, Y., Wu, D., Wu, Q., Xie, F., Xuan, Z., Xue, Y., Yan, C., Yang, X., Yu, B.,  
 Zeng, Y., Zhang, G., Zhang, H., Zhang, H., Zhang, L., Zhang, L., Zhang, M.,  
 Zhang, X., Zhang, X., Zhang, Y., Zhang, Y., Zhang, Z., Zhu, B., Zhu, N.,  
 Yu, J. and Yang, H.

COMMENT

JOURNAL

AUTHORS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Direct Submission  
 Submitted (08-NOV-2002) Human Genomic Center, Institute of  
 Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing  
 100101, P.R.China  
 On Oct 23, 2002 this sequence version replaced gi:13899378.  
 -----Genome Center  
 Center: Beijing Center  
 Center code: Beijing  
 Website: http://hgsc.igtp.ac.cn  
 http://www.genomics.org.cn  
 Contact: hgsc@igtp.ac.cn  
 -----Project Information  
 Center project name: 11% project  
 Center clone name: RP11-1102N22  
 -----Summary Statistics  
 Sequencing vector: pUC18; 100% of reads  
 Chemistry: Dye-terminator; ET 55% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 183208 bases at least Q40  
 Consensus quality: 184666 bases at least Q30  
 Consensus quality: 185280 bases at least Q20  
 Insert size: 181736; sum-of-contigs  
 Quality coverage: 13.39x in Q20 bases; sum-of-contigs

## FEATURES

source

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Location/Qualifiers  
 1. 181736  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="3"  
 /map="3p"  
 /clone="RP11-1102N22"  
 ORIGIN  
 Query Match 63.1%; Score 21; DB 5; Length 181736;  
 Best Local Similarity 51.2%; Pred. No. 5.2e+02;  
 Matches 21; Conservative 7; Mismatches 13; Indels 0; Gaps 0;  
 QY 2 CNCAYGNTGGCCNTCCNCTAYAGGATYAGTGATNAAR 42  
 Db 167657 CTCTCTCAGCCCTTCCCAATGTAAGACCGATGAGTTAA 167617

## RESULT 54

AC018832

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AC018832 185496 bp DNA linear PRI 08-NOV-2002  
 Homo sapiens chromosome 3 clone RP11-48107 map 3p, complete  
 sequence.  
 AC018832  
 AC018832.5 GI:24270675  
 HTG.  
 Homo sapiens (human)  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homiidae; Homo.  
 1 (bases 1 to 185496)  
 Niu, Y., Bao, J., Bao, Q., Bao, W., Bian, X., Cao, T., Chen, C., Chen, J.,









may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

----- Project Information

Center project name: FOVR  
Center clone name: CH240-267C23

----- Summary Statistics

Assembly program: Atlas 3.0;  
Consensus quality: 207919 bases at least Q40  
Consensus quality: 209976 bases at least Q30  
Consensus quality: 211594 bases at least Q20  
Estimated insert size: 218336; sum-of-contigs estimation  
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 19 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 13378: contig of 13378 bp in length  
\* 13379 13428: gap of 50 bp  
\* 13429 28731: contig of 15303 bp in length  
\* 28732 28781: gap of 50 bp  
\* 28782 37340: contig of 8559 bp in length  
\* 37341 37440: gap of unknown length  
\* 37441 40565: contig of 3125 bp in length  
\* 40566 41066: gap of 501 bp  
\* 41067 56282: contig of 15216 bp in length  
\* 56283 56382: gap of unknown length  
\* 56383 58696: contig of 2314 bp in length  
\* 58697 59113: gap of 417 bp  
\* 59114 69258: contig of 10145 bp in length  
\* 69259 69308: gap of 50 bp  
\* 69309 102240: contig of 32932 bp in length  
\* 102241 102290: gap of 50 bp  
\* 102291 105568: contig of 3278 bp in length  
\* 105569 105618: gap of 50 bp  
\* 105619 110100: contig of 4482 bp in length  
\* 110101 110150: gap of 50 bp  
\* 110151 112057: contig of 1907 bp in length  
\* 112058 112107: gap of 50 bp  
\* 112108 139238: contig of 27131 bp in length  
\* 139239 139288: gap of 50 bp  
\* 139289 161152: contig of 21864 bp in length  
\* 161153 161202: gap of 50 bp  
\* 161203 169443: contig of 8247 bp in length  
\* 169444 169493: gap of 50 bp  
\* 169494 173533: contig of 4040 bp in length  
\* 173534 173589: gap of 50 bp  
\* 173590 204051: contig of 30462 bp in length  
\* 204052 204151: gap of unknown length  
\* 204152 206366: contig of 2215 bp in length  
\* 206367 206466: gap of unknown length  
\* 206467 208010: contig of 1544 bp in length  
\* 208011 208110: gap of unknown length  
\* 208111 215523: contig of 7413 bp in length.

# FEATURES source

1. 215523  
/organism="Bos taurus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9913"

/clone="CH240-267C23"  
13379..13428  
/estimated\_length=50  
28732..28781  
/estimated\_length=50  
37341..37440  
/estimated\_length=unknown  
40566..41066  
/estimated\_length=501  
56283..56382  
/estimated\_length=unknown  
58697..59113  
/estimated\_length=417  
69259..69308  
/estimated\_length=50  
102241..102290  
/estimated\_length=50  
105569..105618  
/estimated\_length=50  
110101..110150  
/estimated\_length=50  
112058..112107  
/estimated\_length=50  
139239..139288  
/estimated\_length=50  
161153..161202  
/estimated\_length=50  
169450..169499  
/estimated\_length=50  
173540..173589  
/estimated\_length=50  
204052..204151  
/estimated\_length=unknown  
206367..206466  
/estimated\_length=unknown  
208011..208110  
/estimated\_length=unknown

## ORIGIN

Query Match 63.3%; Score 21; DB 12; Length 215523;  
Best Local Similarity 51.2%; Pred.No. 5.3e+02;  
Matches 21; Conservative 7; Mismatches 13; Indels 0; Gaps 0;  
Qy 2 CNCAYGCGTGGCNCNTCAVAAAGAYTAYGARGTNAAR 42  
Db 85237 CATATGCTTGACCCCTCTTTATATGATGTAATAAG 85277

## RESULT 58

AC154912/c 228141 bp DNA linear ROD 08-MAR-2005  
LOCUS Mus musculus BAC clone RP23-23B10 from 17, complete sequence.  
DEFINITION AC154912  
ACCESSION AC154912 GI:60543593  
VERSION AC154912.3  
KEYWORDS HTG.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 228141)  
Hlafka, M., Kozlowski, A., Bielicki, L. and Haglund, K.  
The sequence of Mus musculus BAC clone RP23-23B10  
Unpublished (2001)  
REFERENCE 2 (bases 1 to 228141)  
WILSON, R. K.  
TITLE Direct Submission  
JOURNAL Submitted (04-JAN-2005) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA  
REFERENCE 3 (bases 1 to 228141)  
WILSON, R. K.  
TITLE Direct Submission  
JOURNAL Submitted (08-FEB-2005) Genome Sequencing Center, 4444 Forest Park





LOCUS	AC109060	246535 bp	DNA	linear	HTG 08-OCT-2002
DEFINITION	Rattus norvegicus clone CH230-9506, *** SEQUENCING IN PROGRESS ***, 3 unordereed pieces.				
ACCESSION	AC109060				
VERSION	AC109060.4	GI:23195511			
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.				
SOURCE	Rattus norvegicus				
ORGANISM	Rattus norvegicus (Norway rat)				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Rattus.				
AUTHORS	1 (bases 1 to 246535)				
	Muzny, D., Marie, Metzker, M., Lee, A., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Frazer, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gegeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshewa, L., Loulseghe, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mathewiney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwankwemeluh, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartabeyn, A., Sisson, I., Sitter, C. D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steelme, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstein, G. and Gibbs, R. A.				
TITLE	Direct Submission				
JOURNAL	Unpublished				
AUTHORS	2 (bases 1 to 246535)				
JOURNAL	Worley, K. C.				
TITLE	Direct Submission				
JOURNAL	Submitted (03-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA				
REFERENCE	3 (bases 1 to 246535)				
AUTHORS	Rat Genome Sequencing Consortium.				
TITLE	Direct Submission				
JOURNAL	Submitted (08-OCT-2002) Human Genome Sequencing Center, Department				

of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA	
On Sep 19, 2002 this sequence version replaced gi:21737852.	
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.	
Center: Baylor College of Medicine	
Center code: BCM	
Web site: http://www.hgsc.bcm.tmc.edu/	
Contact: hgsc-help@bcm.tmc.edu	
Project Information	
Center project name: GPVU	
Center Clone name: CH230-9506	
Summary Statistics	
Assembly program: Phrap; version 0.990329	
Consensus quality: 194634 bases at least Q40	
Consensus quality: 199097 bases at least Q30	
Consensus quality: 202096 bases at least Q20	
Estimated insert size: 216622; sum-of-contigs estimation	
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation	
* NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).	
* NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.	
* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.	
* 1 90577: contig of 90577 bp in length	
* 90578 90677: gap of unknown length	
* 90678 244089: contig of 153412 bp in length	
* 244090 244189: gap of unknown length	
* 244190 246535: contig of 2346 bp in length.	
Location/Qualifiers	
1..246535	/organism="Rattus norvegicus"
	/mol_type="genomic DNA"
	/db_xref="taxon:10116"
	/clone="CH230-9506"
1..1367	/note="wgs_contig"
4513..5520	/note="wgs_contig"
26837..27898	/note="wgs_contig"
90578..90677	/estimated_length=unknown
177840..179232	/note="wgs_contig"
179283..182690	/note="wgs_contig"
244090..244189	/estimated_length=unknown
ORIGIN	
Query Match	63.3%; Score 21; DB 12; Length 246535;
Best Local Similarity	58.3%; Pred No. 5, 5e+02;
Matches	21; Conservative 5; Mismatches 10; Indels 0; Gaps 0;
QY	5 AYCGMTGGCNCNTCCNTAYAARGATYATGARGTNA 40

of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA	
On Sep 19, 2002 this sequence version replaced gi:21737852.	
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.	
Center: Baylor College of Medicine	
Center code: BCM	
Web site: http://www.hgsc.bcm.tmc.edu/	
Contact: hgsc-help@bcm.tmc.edu	
----- Project Information	
Center project name: GPVU	
Center clone name: CH230-9506	
----- Summary Statistics	
Assembly program: Phrap; version 0.990329	
Consensus quality: 194634 bases at least Q40	
Consensus quality: 199097 bases at least Q30	
Consensus quality: 202096 bases at least Q20	
Estimated insert size: 216622; sum-of-contigs estimation	
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation	
-----	
* NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).	
* NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.	
* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.	
* 1 90577: contig of 90577 bp in length	
* 90578 90677: gap of unknown length	
* 90678 244089: contig of 153412 bp in length	
* 244090 244189: gap of unknown length	
* 244190 246535: contig of 2346 bp in length.	
Location/Qualifiers	
1. .246535	
/organism="Rattus norvegicus"	
/mol_type="Genomic DNA"	
/db_xref="taxon:10116"	
/clone="CH230-9506"	
1. .1367	
/note="wgs contig"	
4513. 5520	
/note="wgs contig"	
26837. .27898	
/notes="wgs contig"	
90578. .90677	
/estimated_length=unknown	
177840. .179232	
/note="wgs contig"	
179283. .182690	
/note="wgs contig"	
244090. .244189	
/estimated_length=unknown	
ORIGIN	
Query Match 63.3%; Score 21; DB 12; Length 246535;	
Best Local Similarity 58.3%; Pred. No. 5;Se+02;	
Matches 21; Conservative 5; Mismatches 10; Indels 0; Gaps 0;	
QY 5 AYGCNTGGCCNTCCNTAYARGAYTAYGARGTNA 40	





```

Best Local Similarity 55.0%; Pred. No. 5.9e+02;
Matches 22; Conservative 5; Mismatches 13; Indels 0; Gaps 0;

QY 1 GCNCAYCNCCTGCCNCCNTAYARGAYTAYGARGTNA 40
   |||:|||||:|||||:|||||:|||||:|||||:
Db 58698 GCCCACCACCTGTCCTCCCTCCAGTAGAAGGACTGCGAGGCGA 58659

RESULT 66
CR354544 CR354544 139509 bp DNA linear HTG 05-SEP-2005
LOCUS Danio rerio chromosome 23 clone CH211-273A21, WORKING DRAFT
DEFINITION SEQUENCE, 3 unordered pieces.
ACCESSION CR354544
VERSION CR354544.4 GI:74197629
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 139509)
McLaren, S.
Direct Submission
Submitted (01-SEP-2005) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfisch-help@sanger.ac.uk Clone requests:
http://www.sanger.ac.uk/Projects/D_rerio/faqs.shtml#dataeight
On Sep 6, 2005 this sequence version replaced GI:45561022.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfisch-help@sanger.ac.uk
----- Project Information
Center project name: zk23G9
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 138778 bases at least Q40
Consensus quality: 138874 bases at least Q30
Consensus quality: 138966 bases at least Q20
Insert size: 139309; sum-of-contigs
Insert size: 166839; 1.5% error; agarose-ff
Quality coverage: 8.53x in Q20 bases; sum-of-contigs Quality
coverage: 7.17x in Q20 bases; agarose-ff
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 12730: contig of 12730 bp in length
* 12731 12830: gap of 100 bp
* 12831 43204: contig of 30374 bp in length
* 43205 43304: gap of 100 bp
* 43305 139509: contig of 96205 bp in length.
*
* Location/Qualifiers
* 1. 139509
* /organism="Danio rerio"
* /mol_type="genomic DNA"
* /db_xref="taxon:7955"
* /chromosome="23"
* /clone="CH211-273A21"
* /clone_lib="CHORI-211"
* 1. 12730
* /notes="assembly_fragment:01346
* fragment_chain:1
* clone_end:SP6
* vector_side:left"

misc_feature
12831..43204
/notes="assembly_fragment:01493
fragment_chain:1"
43305..139509
/notes="assembly_fragment:00001
fragment_chain:1"

ORIGIN
Query Match 62.7%; Score 20.8; DB 12; Length 139509;
Best Local Similarity 52.4%; Pred. No. 6.2e+02;
Matches 22; Conservative 6; Mismatches 14; Indels 0; Gaps 0;

QY 1 GCNCAYCNCCTGCCNCCNTAYARGAYTAYGARGTNAAR 42
   |||:|||||:|||||:|||||:|||||:|||||:
Db 55915 GCGTGTCGAAGCCCTCTCTTTCAGGCCCTATGATGTGAAA 55956

RESULT 67
AL662881 AL662881 154959 bp DNA linear ROD 09-FEB-2005
LOCUS Mouse DNA sequence from clone RP23-321C9 on chromosome 11 Contains
DEFINITION part of the gene for a novel protein similar to Tensin Tns,
complete sequence.
ACCESSION AL662881
VERSION AL662881.10 GI:20068665
KEYWORDS HTG; Tensin.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 154959)
White, S.
Direct Submission
Submitted (04-FEB-2005) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
Clone requests: clonerequest@sanger.ac.uk
On Apr 7, 2002 this sequence version replaced GI:19572489.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
-----
Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: vegas@sanger.ac.uk
-----
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one subclone; and the assembly was confirmed by restriction digest,
except on the rare occasion of the clone being a YAC.
Sequence from the Mouse Genome Sequencing Consortium whole genome
shotgun may have been used to confirm this sequence. Sequence data
from the whole genome shotgun alone has only been used where it has
a phred quality of at least 30.
RP23-321C9 is from the RPCI-23 Mouse BAC Library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6.
Location/Qualifiers
1. 154959
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="11"
/clone="RP23-321C9"
/clone_lib="RPCI-23"
join(complement(AL645636.22:38537..38593),

```





```

KEYWORDS
SOURCE      Xenopus tropicalis (Silurana tropicalis)
ORGANISM    Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1 (bases 1 to 165987)
AUTHORS    DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE      Direct Submission
JOURNAL
REFERENCE   2 (bases 1 to 165987)
AUTHORS    DOE Joint Genome Institute.
TITLE      Direct Submission
JOURNAL    Submitted (18-JAN-2004) Production Genomics Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA
            94598-1698, USA
REFERENCE   3 (bases 1 to 165987)
AUTHORS    DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE      Direct Submission
JOURNAL    Submitted (13-APR-2004) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT    On Apr 13, 2004 this sequence version replaced gi:41016131.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www.shgc.stanford.edu
            Quality: Phrap Quality >=40 100% of Sequence;
            Estimated total Number of Errors is 0.
FEATURES             source
                     1..165987
                        /organism="Xenopus tropicalis"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:8364"
                        /clone="CH216-117C13"
ORIGIN
Query Match      62.7%; Score 20.8; DB 11; Length 165987;
Best Local Similarity 52.4%; Pred. No. 6.4e+02;
Matches 22; Conservative 6; Mismatches 14; Indels 0; Gaps 0;

QY      1 GCNAYGCGTGGCNCNTCNCATYARGAVTAYGARGTNAAR 42
      | || || || || || || || || || || || || || || ||
Db      30040 GACCAAGCTTGCTGCTCCAGACAAAGTTATCAGCTGAAA 29999

RESULT 69
AC131779
LOCUS      AC131779 171978 bp DNA linear ROD 28-JAN-2005
DEFINITION Mus musculus BAC clone RP24-336D11 from 8, complete sequence.
ACCESSION  AC131779
VERSION     AC131779.4 GI:56236242
KEYWORDS   HTG.
SOURCE      Mus musculus (house mouse)
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 171978)
AUTHORS    Tomlinson,C., Cotton,M. and Haakenson,W.
TITLE      The sequence of Mus musculus BAC clone RP24-336D11
JOURNAL    Unpublished (2001)
REFERENCE   2 (bases 1 to 171978)
AUTHORS    McPherson,J.D. and Waterston,R.H.
TITLE      Direct Submission
JOURNAL    Submitted (25-AUG-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3 (bases 1 to 171978)
AUTHORS    Wilson,R.K.
TITLE      Direct Submission
JOURNAL    Submitted (28-SEP-2004) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   4 (bases 1 to 171978)
AUTHORS    Wilson,R.K.
TITLE      Direct Submission
JOURNAL    Submitted (03-DEC-2004) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
            5 (bases 1 to 171978)
            Wilson,R.K.
            Direct Submission
            Submitted (28-JAN-2005) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
            On Dec 3, 2004 this sequence version replaced gi:52782655.
            ----- Genome Center
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu
            Contact: submissions@wustl.edu
            ----- Summary Statistics
            -----
            Center project name: M_BB0336D11
            -----
NOTICE
This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.
MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren,
Department of Genetics, Washington University, St. Louis MO. For
additional information about the map position of this sequence, see
http://genome.wustl.edu
SOURCE INFORMATION:
The RPCI-24 BAC Library has been constructed by Pieter de Jong and
coworkers (http://www.chori.org) from male C57BL/6J mouse spleen
and/or brain genomic DNA. The clone and detailed information can be
obtained from Pieter de Jong and coworkers at http://www.chori.org
NEIGHBORING SEQUENCE INFORMATION:
This sequence is the entire insert of the clone.
FEATURES             source
                     1..171978
                        /organism="Mus musculus"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:10090"
                        /chromosome="8"
                        /map="8"
                        /clone="RP24-336D11"
                        /clone_lib="RPCI-24"
                        /notes="Sequence derived from PCR product of project DNA."
misc_feature       117242..117379
unsure             117361
unsure             117376
repeat_region      129522..129523
unsure             /transposon="Bacterial transposon insertion in clone
                        excised here."
unsure             137687..138228
unsure             /notes="Unresolved simple sequence repeat."
misc_feature       138229..138312
unsure             /notes="Sequence derived from PCR product of project DNA."
unsure             140133..140250
unsure             /note="Unresolved simple sequence repeat."
ORIGIN
Query Match      62.7%; Score 20.8; DB 6; Length 171978;
Best Local Similarity 55.0%; Pred. No. 6.4e+02;
Matches 22; Conservative 5; Mismatches 13; Indels 0; Gaps 0;

```



```

misc_feature 7419. 22998
/note="assembly_fragment"
gap 22999. 23098
/estimated_length=unknown
misc_feature 23099. 35627
/note="assembly_fragment"
gap 35628. 35727
/estimated_length=unknown
misc_feature 35728. 55335
/note="assembly_fragment"
gap 55336. 55435
/estimated_length=unknown
misc_feature 55436. 77882
/note="assembly_fragment"
gap 77883. 77982
/estimated_length=unknown
misc_feature 77983. 101387
/note="assembly_fragment"
gap 101388. 101487
/estimated_length=unknown
misc_feature 101488. 138406
/note="assembly_fragment"
gap 138407. 138506
/estimated_length=unknown
misc_feature 138507. 162815
/note="assembly_fragment"
gap 162816. 162915
/estimated_length=unknown
misc_feature 162916. 174929
/note="assembly_fragment
clone_end:T7
vector_side:right"

ORIGIN
Query Match 62.7% Score 20.8; DB 12; Length 174929;
Best Local Similarity 55.0%; Pred. No. 6.4e+02;
Matches 22; Conservative 5; Mismatches 13; Indels 0; Gaps 0;

QY 1 GCNCAVCNTGCCNTCCNTAYAGATYAGARTNA 40
Db 150532 GCACGCGGGCGTCCCATATTTGGACAATGAATTTA 150493

RESULT 71
AC132640
LOCUS Rattus norvegicus clone CH230-339B14, WORKING DRAFT SEQUENCE.
DEFINITION AC132640
ACCESSION AC132640
VERSION 1
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Rattus.
1 (bases 1 to 194709)
Muzny,D,Maric,M, Metzker,M, Lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Cesat,H., Chen,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carrroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escoto,M., Eugene,C., Evans,C.A., Fallis,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebrageorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,

```

```

Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlerbird,P., Jackson,A.,
Jackson,L., Jacob,H., Jiang,L., Johnson,B., Johnson,R., Jolivet,A.,
Karpachy,S., Kelly,S., Khan,Z., Klevan,J., Lewis,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorenaheva,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Parks,K.,
Nwaokemele,O., Okwuonu,G., Olarinpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Shen,H.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shateman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villaseana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczyk,R., Woodden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,X., Zhao,S., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 194709)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (02-SEP-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 194709)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 15, 2002 this sequence version replaced gi:23908481.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: KBQF
Center clone name: CH230-339B14
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 188047 bases at least Q40
Consensus quality: 189447 bases at least Q30
Consensus quality: 190412 bases at least Q20
Estimated insert size: 195068; sum-of-contigs estimation

```

```

-----
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
* NOTE: Estimated insert size may differ from sequence length
  (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
  consists of 1 contigs. Gaps between the contigs
  are represented as runs of N. The order of the pieces
  is believed to be correct as given, however the sizes
  of the gaps between them are based on estimates that have
  provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 194709: contig of 194709 bp in length.
Location/Qualifiers
1. 194709
  /organism="Rattus norvegicus"
  /mol_type="genomic DNA"
  /db_xref="taxon:10116"
  /clone="CH230-339B14"
1. 3065
  /notes="wgs end _extension
clone_end:T7"
7050..7929
  /notes="clone_boundary
clone_end:T7
site:
end_sequence:B2164292"
complement(191365..192203)
  /note="clone_boundary
clone_end:Sp6
site:
end_sequence:B2164293"

FEATURES
source
1. 194709
misc_feature
1. 3065
misc_feature
7050..7929
misc_feature
complement(191365..192203)
misc_feature
clone_end:Sp6
end_sequence:B2164293"

ORIGIN
Query Match 62.7%; Score 20.8; DB 12; Length 194709;
Best Local Similarity 52.4%; Pred. No. 6.6e+02;
Matches 22; Conservative 6; Mismatches 14; Indels 0; Gaps 0;

QY 1 GCNACGTCNTGCCNTCCNTAYARGAYTAYGARGTNAAR 42
|||:||||| |||:|||||:|||||:|||||:
Db 169254 GCCAAGCTGTTGGCCAGTCGACAGACAGAGTAAAG 169295
|||:||||| |||:|||||:|||||:|||||:

RESULT 72
AC126211 195892 bp DNA linear HTG 12-OCT-2002
LOCUS Rattus norvegicus clone CH230-41316, WORKING DRAFT SEQUENCE.
DEFINITION AC126211
AC126211 HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
VERSION Rattus norvegicus (Norway rat)
KEYWORDS SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Muridae; Murinae; Rattus.
Muzny,D,Marie., Metzker,M, Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Faves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebrgeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,

```

```

Gumaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.B., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Loulsegad,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokemele,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Fuazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smaiz,D.,
Speed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steinle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 195892)
Worley,K.C.
Direct Submission
Submitted (04-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 195892)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (12-OCT-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Oct 12, 2002 this sequence version replaced gi:21700469.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GYPP
Center clone name: CH230-41316
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 171920 bases at least Q40
Consensus quality: 173922 bases at least Q30
Consensus quality: 175324 bases at least Q20

```

```

-----
Estimated insert size: 176162; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
  (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).
* NOTE: This is a 'working draft' sequence. It currently
  consists of 1 contigs. Gaps between the contigs
  are represented as runs of N. The order of the pieces
  is believed to be correct as given, however the sizes
  of the gaps between them are based on estimates that have
  been provided by the submitter.
* This sequence will be replaced
  as soon as it is available and the accession number will
  be preserved.
* By the finished sequence as soon as it is available and
  the accession number will be preserved.
-----
FEATURES
source
    1..195892
    /organism="Rattus norvegicus"
    /mol_type="genomic DNA"
    /db_xref="taxon:10116"
    /clone="CH230-41316"
misc_feature
    1..1313
    /note="wgs_end_extension
clone_end:T7"
misc_feature
    1815..2881
    /note="wgs_end_extension
clone_end:T7"
misc_feature
    complement(11158..12220)
    /note="clone_boundary
clone_end:T7
site:Mbol
end_sequence:RXBERS1TVJ"
misc_feature
    194721..195289
    /note="clone_boundary
clone_end:Sp6
site:Mbol
end_sequence:RXBERS1TVJ"
ORIGIN
Query Match      62.7%; Score 20.8; DB 12; Length 195892;
Best Local Similarity 59.5%; Pred. No. 6.6e+02;
Matches 22; Conservative 4; Mismatches 11; Indels 0; Gaps 0;
QY 1 GCNCAVCGTGGCCNCCNTAAYAGAYTAYGARG 37
Db 109270 GTGCATTTCGGCCCTTCTTACAGATAGAGG 109306
-----
RESULT 73
CT573319/c
LOCUS CT573319 210916 bp DNA linear HTG 30-JAN-2006
DEFINITION Danio rerio chromosome 23 clone DKEY-3K23, *** SEQUENCING IN
PROGRESS ***, 24 unordered pieces.
ACCESSION CT573319
VERSION CT573319.2 GI:86197632
KEYWORDS HTG; HTGS PHASE1.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 210916)
Sims,S.
Direct Submission
Submitted (29-JAN-2006) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfish-help@sanger.ac.uk Clone requests:
http://www.sanger.ac.uk/Projects/D\_rerio/faq.shtmldataeight
On Jan 30, 2006 this sequence version replaced gi:85857250.
-----
FEATURES
source
    1..210916
    /organism="Danio rerio"
    /mol_type="genomic DNA"
-----
Contact: zfish-help@sanger.ac.uk
-----
Project Information
Center project name: zK3K23
-----
Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 198408 bases at least Q40
Consensus quality: 200547 bases at least Q30
Consensus quality: 202510 bases at least Q20
Insert size: 208616; sum-of-contigs
Insert size: 194803; 3.0% error; agarose-fp
Quality coverage: 3.97x in Q20 bases; sum-of-contigs Quality
coverage: 4.29x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
  consists of 24 contigs. The true order of the pieces
  is not known and their order in this sequence record is
  arbitrary. Gaps between the contigs are represented as
  runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
  as soon as it is available and the accession number will
  be preserved.
-----
1 5453: contig of 5453 bp in length
* 5454: gap of 100 bp
5553: contig of 3825 bp in length
9378: contig of 100 bp
9478: gap of 100 bp
19783: contig of 10305 bp in length
19883: gap of 100 bp
39109: contig of 19226 bp in length
39209: gap of 100 bp
62403: contig of 23194 bp in length
62503: gap of 100 bp
73792: contig of 11289 bp in length
73892: gap of 100 bp
78875: contig of 4983 bp in length
78975: gap of 100 bp
87638: contig of 8663 bp in length
87738: gap of 100 bp
87639: contig of 5389 bp in length
93127: contig of 100 bp
93227: gap of 100 bp
95248: contig of 2021 bp in length
95348: gap of 100 bp
95349: contig of 4536 bp in length
99884: gap of 100 bp
99885: contig of 4735 bp in length
104719: contig of 100 bp
104720: gap of 100 bp
116506: contig of 11687 bp in length
116507: gap of 100 bp
116607: contig of 14034 bp in length
130640: gap of 100 bp
130641: contig of 8770 bp in length
139510: gap of 100 bp
139511: contig of 4788 bp in length
144398: gap of 100 bp
144399: contig of 2152 bp in length
146650: gap of 100 bp
146651: contig of 3231 bp in length
149981: gap of 100 bp
150081: contig of 100 bp
158194: contig of 8113 bp in length
158195: gap of 100 bp
158294: contig of 6953 bp in length
165247: gap of 100 bp
165248: contig of 10632 bp in length
175979: gap of 100 bp
176079: gap of 100 bp
176080: contig of 3605 bp in length
179684: gap of 100 bp
179685: contig of 5765 bp in length
185549: gap of 100 bp
185550: contig of 25267 bp in length.
185650: gap of 100 bp
210916: contig of 25267 bp in length.
-----
Location/Qualifiers
1..210916
/organism="Danio rerio"
/mol_type="genomic DNA"
-----

```

```
/db_xref="taxon:7955"
/chromosomes="23"
/clone="DKEY-3K23"
/clone_lib="DanioKey"
1. .5453
/notes="assembly fragment:00258
fragment chain:1"
5554. .9378
/notes="assembly fragment:00056
fragment chain:1"
9479. .13783
/notes="assembly fragment:00422
fragment chain:1"
19884. .39109
/notes="assembly fragment:00771
fragment chain:1"
39210. .52403
/notes="assembly fragment:00877
fragment chain:1"
62504. .73792
/notes="assembly fragment:00605
fragment chain:1"
73993. .78875
/notes="assembly fragment:00167
fragment chain:1"
78976. .87638
/notes="assembly fragment:00371
fragment chain:2"
87739. .93127
/notes="assembly fragment:00195
fragment chain:2"
93228. .95248
/notes="assembly fragment:00010
fragment chain:2"
95349. .99884
/notes="assembly fragment:00094
fragment chain:2"
99985. .104719
/notes="assembly fragment:00075
fragment chain:2"
104820. .116506
/notes="assembly fragment:00538
fragment chain:2"
116607. .130640
/notes="assembly fragment:00678
fragment chain:2"
130741. .139510
/notes="assembly fragment:00225
fragment chain:3"
139611. .144398
/notes="assembly fragment:00113
fragment chain:3"
144499. .146650
/notes="assembly fragment:00015
fragment chain:3"
146751. .149981
/notes="assembly fragment:00028
fragment chain:4"
150082. .158194
/notes="assembly fragment:00329
fragment chain:4"
158295. .165247
/notes="assembly fragment:00140
fragment chain:5"
165348. .175979
/notes="assembly fragment:00477
fragment chain:5"
176080. .179684
/notes="assembly fragment:00038"
179785. .185549
/notes="assembly fragment:00293.0"
185650. .210916
/notes="assembly fragment:01006"
```

ORIGIN

Query Match 62.7%; Score 20.8; DB 12; Length 210916;  
Best Local Similarity 52.4%; Pred. No. 6.7e+02;  
Matches 22; Conservative 6; Mismatches 14; Indels 0; Gaps 0;

QY 1 GCNCAYGCTGGCCNTCNCNTAYAAAGAYTAYAGGTNAAR 42  
DB 180085 GCGTGTGCAAGCCCTCTTCTTTCAAGGCTATGATGTGAAA 180044

RESULT 74  
AC096459/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AC096459 212324 bp DNA linear HTG 10-MAY-2003  
Rattus norvegicus clone CH230-147D4, WORKING DRAFT SEQUENCE.  
AC096459.6 GI:30521400  
HTG: HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.  
Rattus norvegicus (Norway rat)  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridea; Muridae; Murinae; Rattus.  
1 (bases 1 to 212324)  
Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,  
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,  
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,  
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,  
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,  
Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A.,  
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,  
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
Davila, M. L., Davis, S., Davy-Carroll, L., De Anda, C., Dederich, D.,  
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,  
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,  
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,  
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
Georgiev, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,  
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,  
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hoques, M.,  
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,  
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,  
Karpachy, S., Kelly, S., Khan, Z., King, L., Kovar, C.,  
Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
Lorensuewa, L., Loulseged, H., Lozado, R. J., Lu, X., Ma, J.,  
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,  
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,  
Mawhiney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,  
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,  
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,  
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,  
Nwaokeme, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,  
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,  
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., L.,  
Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,  
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,  
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,  
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D.,  
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,  
Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,  
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,  
Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J.,  
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,  
Williams, G., Willson, R., Wlarczyk, R., Wooden, H., Worley, K.,  
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
Yu, F., Zhang, J., Zhou, X., Zhou, X., Zhao, S., Dunn, D., von  
Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,  
Weinstock, G. and Gibbs, R. A.

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Direct Submission
Unpublished
2 (bases 1 to 212324)
Worley, K.C.
Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 212324)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 10, 2003 this sequence version replaced gi:24942666.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GFHW
Center clone name: CH230-147D4
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 205118 bases at least Q40
Consensus quality: 206969 bases at least Q30
Consensus quality: 208110 bases at least Q20
Estimated insert size: 21332; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 212324: contig of 212324 bp in length.
FEATURES
source
1. .212324
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-147D4"

ORIGIN
Query Match 62.78; Score 20.8; DB 12; Length 212324;
Best Local Similarity 52.44; Pred. No. 6.7e+02;
Matches 22; Conservative 6; Mismatches 14; Indels 0; Gaps 0;

QY 1 GCNCAYCNGTGGCCNTCNCNTAYARGATYATGAGTNAAR 42
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 192354 GCACATGTGTCTCATACATTTACACACAACTGAATTTAAA 192313
|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 75

```

AC137342/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AC137342 219429 bp DNA linear HTG 20-NOV-2002  
Rattus norvegicus clone CH230-unknown, WORKING DRAFT SEQUENCE.  
AC137342  
AC137342.1 GI:25138433  
HTG; HTGS PHASE2; HTGS DRAFT; HTGS\_FULLTOP.  
Rattus norvegicus (Norway rat)  
Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Rattus;  
1 (bases 1 to 219429)

REFERENCE  
AUTHORS

Muzny, D., Maric, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J.,  
Allen, C., Allen, H., Alebrooks, S., Anin, A., Anguiano, D.,  
Aryalabechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,  
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,  
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,  
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,  
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,  
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,  
DeGado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,  
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,  
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,  
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
Fraser, C.M., Gabisi, A., Gant, R., Garcia, A., Garner, T., Garza, W.,  
Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,  
Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,  
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
Hollins, B., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,  
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,  
Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,  
Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
Lorensueta, L., Louised, H., Lozano, R.J., Lu, X., Ma, J.,  
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,  
Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,  
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,  
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,  
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,  
Nwankweme, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parke, K.,  
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,  
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,  
Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,  
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,  
Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,  
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D.,  
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,  
Steinle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,  
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,  
Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,  
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,  
Williams, G., Willison, R., Wleczyk, R., Wooden, H., Worley, K.,  
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
Yu, F., Zhang, J., Zhou, X., Zhou, X., Zhao, S., Dunn, D., von  
Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,  
Weinstock, G. and Gibbs, R.A.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Unpublished  
2 (bases 1 to 219429)  
Rat Genome Sequencing Consortium.  
Direct Submission

Submitted (20-NOV-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
The sequence in this assembly is a combination of BAC based reads  
and whole genome shotgun sequencing reads assembled using Atlas  
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described  
in the feature table below represents a scaffold in the Atlas

assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

#### ----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

----- Project Information

Center project name: K2R3K

Center clone name: CH230-unknown

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 200012 bases at least Q40

Consensus quality: 202466 bases at least Q30

Consensus quality: 204274 bases at least Q20

Estimated insert size: 205118; sum-of-contigs estimation

Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

- \* NOTE: Estimated insert size may differ from sequence length (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).
- \* NOTE: This is a 'working draft' sequence. It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.
- \* This sequence will be replaced
- \* by the finished sequence as soon as it is available and the accession number will be preserved.
- \* 1 219429: Contig of 219429 bp in length.

#### FEATURES

##### source

```
1. 219429
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-unknown"
218116..219429
/notes="wgs_contig"
```

##### misc\_feature

##### ORIGIN

```
Query Match      62.7%; Score 20.8; DB 12; Length 219429;
Best Local Similarity 52.4%; Pred. No. 6.7e+02;
Matches 22; Conservative 6; Mismatches 14; Indels 0; Gaps 0;
```

```
QY 1 GCNCAYGCNTGCCNTCNCNTAYARGAYTAYGARGTNAAR 42
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 191598 GCACATGTGTGTCACATACCTTACACACAAATGATTTAA 191557
```

#### RESULT 76

##### AC108634

##### LOCUS

```
DEFINITION Rattus norvegicus clone CH230-249H12, *** SEQUENCING IN PROGRESS
***, 5 unordered pieces.
```

##### AC108634

AC108634.4 GI:23195538

HTG; HTGS PHASE1; HTGS DRAFT; HTGS\_ENRICHED.

Rattus norvegicus (Norway rat)

##### ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Muridae; Murinae; Rattus.

1 (bases 1 to 232234)

Muzny, D., Marie, M., Metzger, M., Lee, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsebrook, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,

Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Center, A., Chacko, J., Chavez, D., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geek, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, C., Hernandez, J., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hognes, M., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hognes, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensu, L., Loulseghe, H., Lozano, R., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokemele, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, P., Poindexter, A., Popovic, D., Primus, E., Pu, L., Puzo, M., Quirroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajz, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

#### Direct Submission

Unpublished

2 (bases 1 to 232234)

Worley, K.C.

#### Direct Submission

Submitted (31-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 232234)

Rat Genome Sequencing Consortium.

#### Direct Submission

Submitted (19-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Sep 19, 2002 this sequence version replaced gi:21737739.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Alas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). As a result, the sequence may extend beyond the ends of the clone and there may be contigs that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information







**TITLE**  
JOURNAL

**COMMENT**

Direct Submission  
Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On May 10, 2003 this sequence version replaced gi:23096320. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: http://www.hgsc.bcm.tmc.edu/  
Contact: hgsc-help@bcm.tmc.edu

----- Project Information  
Center project name: GFWD

----- Summary Statistics  
Center clone name: CH230-103A13

----- Consensus quality: 192971 bases at least Q40

----- Consensus quality: 19252 bases at least Q30

----- Consensus quality: 203322 bases at least Q20

----- Estimated insert size: 207485; sum-of-contigs estimation

----- Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

-----

NOTE: Estimated insert size may differ from sequence length

NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 236414: contig of 236414 bp in length

236515: gap of unknown length

238065: contig of 1350 bp in length

238165: gap of unknown length

239373: contig of 1208 bp in length

239473: gap of unknown length

241557: contig of 2085 bp in length.

Location/Qualifiers

1. 241557

/organism="Rattus norvegicus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10116"

/clone="CH230-103A13"

1. 2015

/note="wgs contig"

114280. 116396

/note="wgs contig"

166435. 168513

/note="wgs contig"

169489. 171488

/note="wgs contig"

236415. 236514

/estimated\_length=unknown

238065. 238164

/estimated\_length=unknown

239373. 239472

/estimated\_length=unknown

**FEATURES**  
source

misc\_feature

misc\_feature

misc\_feature

misc\_feature

gap

gap

gap

**ORIGIN**

Query Match 62.7%; Score 20.8; DB 12; Length 241557;

Best Local Similarity 59.5%; Pred. No. 6.8e+02;  
Matches 22; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

Oy 1 GCNAYCNGTGGCCNCCNTAYARGAYTAYGARG 37

Db 45208 GTGCATTGCTTTTCTTCTTACAAAGATAGGAGG 45244

RESULT 79

AC134740

LOCUS

DEFINITION

Rattus norvegicus clone CH230-81D7, WORKING DRAFT SEQUENCE, 2

unordered pieces.

AC134740.2 GI:25138728

HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.

KEYWORDS

SOURCE

ORGANISM

Rattus norvegicus

(Norway rat)

REFERENCE

AUTHORS

1 (bases 1 to 243005)

Muzny, D. Marie, Metzker, M. Lee, Abramson, S., Adams, C., Alder, J.,

Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,

Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,

Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,

Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,

Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,

Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,

Chacko, J., Chavez, D., Chen, G., Chen, Y., Chen, Z., Chu, J.,

Cleveland, C., Cockrell, R., Cox, C., Coyne, M., Cree, A., D'Souza, L.,

Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,

Dalgaard, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,

Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,

Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,

Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,

Fraser, C.M., Gabis, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,

Georgiadis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,

Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,

Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,

Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,

Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,

Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,

Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,

Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,

Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,

Lorensu, L., Loulseg, H., Lozano, R.J., Lu, X., Ma, J.,

Mareshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,

Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,

Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,

Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,

Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,

Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,

Nwako, L., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,

Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,

Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,

Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,

Reilly, B., Reilly, W., Ren, Y., Reuter, M., Richards, S., Riggs, F.,

Rives, C., Rodkey, T., Rojars, A., Rose, M., Rose, R., Ruiz, S.J.,

Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,

Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D.,

Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,

Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,

Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,

Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J.,

Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,

Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K.,

Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, V.,

Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von

Niederhauser, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,

Neistack, G. and Gibbs, R.A.

Direct Submission

Unpublished

TITLE

JOURNAL

```

REFERENCE
AUTHORS      2 (bases 1 to 243005)
TITLE        Rat Genome Sequencing Consortium.
JOURNAL      Direct Submission
Submitted (30-SEP-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

REFERENCE
AUTHORS      3 (bases 1 to 243005)
TITLE        Rat Genome Sequencing Consortium.
JOURNAL      Direct Submission
Submitted (20-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

COMMENT
On Nov 20, 2002 this sequence version replaced gi:23346744.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: KCVN
Center clone name: CH230-81D7
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 229507 bases at least Q40
Consensus quality: 231935 bases at least Q30
Consensus quality: 233508 bases at least Q20
Estimated insert size: 237697; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 239920: contig of 239920 bp in length
* 239921 240020: gap of unknown length
* 240021 243005: contig of 2985 bp in length.

FEATURES
source
    Location/Qualifiers
        1. 243005
            /organism="Rattus norvegicus"
            /mol_type="genomic DNA"
            /db_xref="taxon:10116"
            /clone="CH230-81D7"
            /complement(239825..239463)
            /notes="clone boundary
            site_end:T7
            end sequence: BH349441"
            239921..240020
            /estimated_length=unknown

misc_feature
    complement(239825..239463)
    site_end:T7

gap
    end sequence: BH349441"
    239921..240020
    /estimated_length=unknown

ORIGIN
Query Match      62.7%; Score 20.8; DB 12; Length 243005;
Best Local Similarity 52.4%; Pred. No. 6.8e+02;
Matches 22; Conservative 6; Mismatches 14; Indels 0; Gaps 0;

1 GCNCAYGCGTGGCCNTCNCNTAYARGAYTAYGARGTNAAR 42
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121437 GCACATGTGTGTCCTACATCAACAACAATGAATTAAA 121478

RESULT 80
AC115384
LOCUS
DEFINITION
    Rattus norvegicus clone CH230-83H13, *** SEQUENCING IN PROGRESS
    ***, 3 unordered pieces.
AC115384.4 GI:24818945
VERSION
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
KEYWORDS
Rattus norvegicus (Norway rat)
SOURCE
ORGANISM
    Rattus norvegicus
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
    Sciurognathi; Muridea; Muridae; Murinae; Rattus.
    1 (bases 1 to 287291)
Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hoques, M.,
Hollins, B., Howell, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenshewa, L., Louisedge, H., Lozado, R. J., Lu, X., Ma, J.,
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwaokeme, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,
Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, X., Zhao, S., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,
Weinstock, G. and Gibbs, R. A.
Direct Submission
Unpublished
2 (bases 1 to 287291)
Worley, K. C.
Direct Submission
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE

```

## JOURNAL

Submitted (19-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 287291)  
 Rat Genome Sequencing Consortium.  
 Direct Submission

REFERENCE  
AUTHORS

## TITLE

Submitted (09-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

## COMMENT

On Nov 9, 2002 this sequence version replaced gi:23264550.  
 The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: http://www.hgsc.bcm.tmc.edu/  
 Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GMPN  
 Center clone name: CH230-83H13  
 ----- Summary Statistics

Assembly program: Phrap; version 0.990329  
 Consensus quality: 236597 bases at least Q40  
 Consensus quality: 240612 bases at least Q30  
 Consensus quality: 243338 bases at least Q20  
 Estimated insert size: 245299; sum-of-contigs estimation  
 Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

-----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 3 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1 274388: contig of 274388 bp in length  
 \* 274389 274488: gap of unknown length  
 \* 274489 285679: contig of 11191 bp in length  
 \* 285680 285779: gap of unknown length  
 \* 285780 287291: contig of 1512 bp in length.

## FEATURES

## source

1. .287291  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-83H13"

## misc\_feature

1. .1457

## misc\_feature

4592..5617

## misc\_feature

6985..7843

## misc\_feature

clone\_end:T7

## misc\_feature

clone\_end:T7

## misc\_feature

clone\_end:T7

## misc\_feature

clone\_end:T7

## misc\_feature

clone\_end:T7

## site:ECORI

end\_sequence:BH273106"

## misc\_feature

271675..274388

## gap

/note="wgs\_end\_extension"

## misc\_feature

274389..274488

## misc\_feature

/estimated\_length=unknown

## misc\_feature

274489..275788

## misc\_feature

/note="wgs\_end\_extension"

## gap

285680..285779

## gap

/estimated\_length=unknown

## ORIGIN

Query Match 62.7%; Score 20.8; DB 12; Length 287291;

Best Local Similarity 52.4%; Pred. No. 7.1e+02;

Matches 22; Conservative 6; Mismatches 14; Indels 0; Gaps 0;

Qy 1 GCNCAYCCTGGCCCTCCTCTGACAGGATTACCATCTCCAA 156915

Db 156874 GTGCACAGTTGGCCATCTCTGACAGGATTACCATCTCCAA 156915

## RESULT 81

## AC110863

## LOCUS

## DEFINITION

AC110863 319548 bp DNA linear HTG 13-NOV-2002

Rattus norvegicus clone CH230-135F12, \*\*\* SEQUENCING IN PROGRESS

\*\*\*, 13 unordered pieces.

## ACCESSION

AC110863 GI:24941814

## VERSION

HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.

## KEYWORDS

Rattus norvegicus (Norway rat)

## SOURCE

Rattus norvegicus

## ORGANISM

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muroidae; Muridae; Murinae; Rattus.

1 (bases 1 to 319548)

Muzny,D,Marie, Metzker,M, Lee, S., Abramson, S., Adams, C., Alder, J.,

Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,

Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,

Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,

Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,

Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,

Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Center, A.,

Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,

Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,

Dalgaard, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,

Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K.,

Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,

Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,

Fraser, C.M., Gabis, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,

Georgiev, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,

Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,

Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,

Hernandez, R., Hines, S., Hladun, S., L., Hodgson, A., Hogues, M.,

Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,

Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,

Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,

Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,

Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,

Lorensuwha, L., Lounsbury, H., Lozano, R. J., Lu, X., Ma, J.,

Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,

Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,

Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,

Milosavljevic, A., Miner, G., Minja, B., Montemayor, J., Moore, S.,

Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,

Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,

Nwaokeme, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,

Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,

Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,  
 Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,  
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
 Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,  
 Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,  
 Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D.,  
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,  
 Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,  
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,  
 Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J.,  
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,  
 Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K.,  
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von  
 Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,  
 Weinstock, G. and Gibbs, R.A.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 319548)  
 Worley, K.C.  
 Direct Submission  
 Submitted (16-FEB-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 319548)  
 Rat Genome Sequencing Consortium.  
 Direct Submission  
 Submitted (13-NOV-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Nov 13, 2002 this sequence version replaced gi:23816394.  
 The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described  
 in the feature table below represents a scaffold in the Atlas  
 assembly (a 'contig-scaffold'). Within each contig-scaffold,  
 individual sequence contigs are ordered and oriented, and separated  
 by sized gaps filled with Ns to the estimated size. The sequence  
 may extend beyond the ends of the clone and there may be sequence  
 contigs within a contig-scaffold that consist entirely of whole  
 genome shotgun sequence reads. Both end sequences and whole genome  
 shotgun sequence only contigs will be indicated in the feature  
 table.

----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
 ----- Project Information  
 Center project name: GSQR  
 Center clone name: CH230-135F12  
 ----- Summary Statistics  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 258869 bases at least Q40  
 Consensus quality: 263450 bases at least Q30  
 Consensus quality: 267288 bases at least Q20  
 Estimated insert size: 267899; sum-of-contigs estimation  
 Quality coverage: 5x in Q20 bases; sum-of-contigs estimation  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html))  
 \* NOTE: This sequence may represent more than one clone.  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 13 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 13284: contig of 13284 bp in length  
 \* 13285 13384: gap of unknown length  
 \* 13385 33286: contig of 19902 bp in length

FEATURES	Location/Qualifiers
source	1..319548
	organism="Rattus norvegicus"
	/mol_type="genomic DNA"
	/db_xref="taxon:10116"
	/clone="CH230-135F12"
gap	13285..13384
	/estimated_length=unknown
misc_feature	24923..26466
	/notes="wgs contig"
gap	33287..33386
	/estimated_length=unknown
misc_feature	33387..34475
	/note="wgs contig"
gap	36574..36673
	/estimated_length=unknown
gap	50543..50642
	/estimated_length=unknown
misc_feature	50643..51978
	/note="wgs contig"
misc_feature	complement(219445..220297)
	/note="clone_boundary"
	clone_end:T7
	site:
	end sequence:RAWL30TUB"
misc_feature	276518..278526
	/note="wgs_end_extension"
gap	278527..278626
	/estimated_length=unknown
misc_feature	281570..282956
	/note="wgs_end_extension"
gap	289356..289455
	/estimated_length=unknown
gap	299609..299708
	/estimated_length=unknown
misc_feature	299709..301747
	/note="wgs_end_extension"
gap	310177..310276
	/estimated_length=unknown
gap	311614..311713
	/estimated_length=unknown
gap	313059..313158
	/estimated_length=unknown
gap	315146..315245
	/estimated_length=unknown
gap	317899..317998
	/estimated_length=unknown

ORIGIN

Query Match 62.7%; Score 20.8; DB 12; Length 319548;  
 Best Local Similarity 52.4%; Pred. No. 7.3e+02;  
 Matches 22; Conservative 6; Mismatches 14; Indels 0; Gaps 0;

Qy 1 GCNCAYGNTGGCCNTCCNTAYAAARGAYTAYGARGTNAAR 42  
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
 Db 191152 GCACATGTGTCCATACACTTACCAACAATGAATTAAA 191193

RESULT 82  
 AX492783/c AX492783 349980 bp DNA linear PAT 26-SEP-2002  
 LOCUS Sequence 1101 from Patent EP127152.  
 DEFINITION  
 ACCESSION AX492783  
 VERSION AX492783.1 GI:23339466

KEYWORDS Bifidobacterium longum biovar Longum  
 SOURCE Bifidobacterium longum biovar Longum  
 ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;  
 Bifidobacteriaceae; Bifidobacterium.

REFERENCE 1  
 AUTHORS None.  
 TITLE Bacterial strain and genome of bifidobacterium  
 JOURNAL Patent: EP 127152-A 1101 31-JUL-2002;  
 SOCIETE des Produits Nestle S.A. (CH)

FEATURES  
 source 1..349980  
 /organism="Bifidobacterium longum biovar Longum"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:1679"  
 /note="1106 seq in place of 1099 because seq 1 (2.256.638  
 bases splitted  
 in 7 more sequences.  
 seq 0001: from 0.000.001 to 0.349.980 length: 349980  
 seq 1100: from 0.300.001 to 0.649.980 length: 349980  
 seq 1101: from 0.600.001 to 0.949.980 length: 349980  
 seq 1102: from 0.900.001 to 1.249.980 length: 349980  
 seq 1103: from 1.200.001 to 1.549.980 length: 349980  
 seq 1104: from 1.500.001 to 1.849.980 length: 349980  
 seq 1105: from 1.800.001 to 2.149.980 length: 349980  
 seq 1106: from 2.100.001 to 2.256.638 length: 156638"

## ORIGIN

Query Match 62.7%; Score 20.8; DB 2; Length 349980;  
 Best Local Similarity 55.0%; Pred. No. 7.3e+02;  
 Matches 22; Conservative 5; Mismatches 13; Indels 0; Gaps 0;

Qy 1 GCNCAYGNTGGCCNTCCNTAYAAARGAYTAYGARGTNA 40  
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:  
 Db 129737 GCCACCACTGTCTCTCCAGTAGAGGACTGCGAGCGA 129698

RESULT 83  
 AX553950/c AX553950 349980 bp DNA linear PAT 27-NOV-2002  
 LOCUS Sequence 4 from Patent WO02074798.  
 DEFINITION  
 ACCESSION AX553950  
 VERSION AX553950.1 GI:25897903

KEYWORDS Bifidobacterium longum biovar Longum  
 SOURCE Bifidobacterium longum biovar Longum  
 ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;  
 Bifidobacteriaceae; Bifidobacterium.

REFERENCE 1  
 AUTHORS Arigoni, F., Delley, M., Mollet, B., Pridmore, R.D., Schell, M.A.,  
 Pohl, T.G. and Zwhlen, M.C.  
 TITLE The genome of a bifidobacterium  
 JOURNAL Patent: WO 02074798-A 4 26-SEP-2002;  
 SOCIETE DES PRODUITS NESTLE S.A. (CH)

FEATURES  
 source 1..349980  
 /organism="Bifidobacterium longum biovar Longum"

/mol\_type="unassigned DNA"  
 /db\_xref="taxon:1679"  
 /note="seq 1: original length 2256646 splitted into  
 seq 1: from 0.000.001 to 0.349.980  
 seq 3: from 0.300.001 to 0.649.980  
 seq 4: from 0.600.001 to 0.949.980  
 seq 5: from 0.900.001 to 1.249.980  
 seq 6: from 1.200.001 to 1.549.980  
 seq 7: from 1.500.001 to 1.849.980  
 seq 8: from 1.800.001 to 2.149.980  
 seq 9: from 2.100.001 to 2.256.646"

## ORIGIN

Query Match 62.7%; Score 20.8; DB 2; Length 349980;  
 Best Local Similarity 55.0%; Pred. No. 7.3e+02;  
 Matches 22; Conservative 5; Mismatches 13; Indels 0; Gaps 0;

Qy 1 GCNCAYGNTGGCCNTCCNTAYAAARGAYTAYGARGTNA 40  
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:  
 Db 129740 GCCACCACTGTCTCTCCAGTAGAGGACTGCGAGCGA 129701

## RESULT 84

BV639777  
 LOCUS S217P60461FC10.T0 Noemie Pan troglodytes troglodytes STS genomic,  
 DEFINITION sequence tagged site.  
 ACCESSION BV639777  
 VERSION BV639777.1 GI:62667747

KEYWORDS STS.  
 SOURCE Pan troglodytes troglodytes  
 ORGANISM Pan troglodytes troglodytes  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homnidae; Pan.

## REFERENCE

1 (bases 1 to 770)  
 Mikkelsen, T.S., Hillier, W.L., Eichler, E.E., Zody, M.C. and  
 Jaffe, D.B.

TITLE Initial Sequence of the Chimpanzee Genome and Comparison with the  
 HUMAN Genome  
 JOURNAL Unpublished (2005)  
 COMMENT

Contact: Michael C. Zody  
 Broad Institute of MIT and Harvard  
 320 Charles Street, Cambridge, MA 02141, USA  
 Tel: 6172580933  
 Fax: 6172580903

Email: mczody@broad.mit.edu  
 Primer A: No sequence submitted  
 Primer B: No sequence submitted  
 STS size: 770

Protocol:  
 23,021,928 chimpanzee whole genome shotgun reads were aligned to  
 the Human genome NCBI  
 Build 34 (hg16, July 2003). Chimp WGS reads were from 9 donors,  
 including Clint (Pan  
 troglodytes verus), 3 other Pan troglodytes verus chimps  
 (Donald, Karlén, Yvonne), 3 Pan  
 troglodytes troglodytes chimps (Noemie, Masuku, Clara) and 2 chimps  
 of unknown origin  
 (Gon, Unknown Chimp). Common names: Pan troglodytes verus is the  
 western chimp and Pan  
 troglodytes troglodytes is the central chimp. To be included in  
 chimpanzee SNP discovery, a  
 read must be at least 500bp in length, at least 50% of its base  
 calls must have Phred  
 score >= 20, at least 30% of its base calls must satisfy  
 SNQS(30,25) (single strand NQS, the  
 base in question has Phred score >= 30, the surrounding 10 bases in  
 the read have Phred  
 score >= 25), and the read must have at least 200 bp SNQS(30,25)  
 bases. Reads not uniquely  
 placed in the genome and read pairs whose two ends were not







```
Db          9197 ATGCTGACAATCACCTAATAAGATAATAGTAAG 9234
||||| | | | | | | | | | | | | | | | | | | | | |
AC100253    44308 bp      DNA      linear      HTG 22-NOV-2001
LOCUS       Mus musculus clone RP23-66A2, LOW-PASS SEQUENCE SAMPLING.
DEFINITION  AC100253
ACCESSION   AC100253
VERSION     AC100253.1 GI:17047619
KEYWORDS    HTG: HTGS PHASE0.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidae; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 44308)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE       Mus musculus, clone RP23-66A2
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 44308)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,I., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., Larocque,K.,
            Lamazares,R., Landers,T., Lehocsky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrum,J.,
            Menus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Vriel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
            Direct Submission
            Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ----- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ----- Project Information
            Center project name: L14164
            Center clone name: 66_A_2
            -----
            * NOTE: This record contains 54 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *
            * 1 715: contig of 715 bp in length
            * 716 815: gap of 100 bp
            * 816 1521: contig of 706 bp in length
            * 1522 1621: gap of 100 bp
            *
            * 1622 2351: contig of 730 bp in length
            * 2352 2451: gap of 100 bp
            * 2452 3161: contig of 710 bp in length
            * 3162 3261: gap of 100 bp
            * 3262 3992: contig of 731 bp in length
            * 3993 4093: gap of 100 bp
            * 4093 4826: contig of 734 bp in length
            * 4827 4927: gap of 100 bp
            * 4927 5648: contig of 721 bp in length
            * 5648 5748: gap of 100 bp
            * 5748 6568: gap of 100 bp
            * 6569 7314: contig of 746 bp in length
            * 7315 8125: contig of 711 bp in length
            * 8126 8226: gap of 100 bp
            * 8226 8926: contig of 701 bp in length
            * 8927 9027: gap of 100 bp
            * 9027 9573: contig of 726 bp in length
            * 9573 10573: gap of 100 bp
            * 10573 11390: contig of 720 bp in length
            * 11390 11490: gap of 100 bp
            * 11490 12215: contig of 718 bp in length
            * 12215 12315: gap of 100 bp
            * 12315 13066: contig of 751 bp in length
            * 13066 13166: gap of 100 bp
            * 13166 13876: contig of 710 bp in length
            * 13876 14703: gap of 100 bp
            * 14703 14803: contig of 727 bp in length
            * 14803 15533: gap of 100 bp
            * 15533 16369: contig of 736 bp in length
            * 16369 17174: gap of 100 bp
            * 17174 17241: contig of 705 bp in length
            * 17241 17983: gap of 100 bp
            * 17983 18083: contig of 709 bp in length
            * 18083 18798: gap of 100 bp
            * 18798 18988: contig of 715 bp in length
            * 18988 19623: gap of 100 bp
            * 19623 19723: contig of 725 bp in length
            * 19723 20541: gap of 100 bp
            * 20541 21281: contig of 718 bp in length
            * 21281 21381: gap of 100 bp
            * 21381 22118: contig of 740 bp in length
            * 22118 22218: gap of 100 bp
            * 22218 22908: contig of 737 bp in length
            * 22908 23723: gap of 100 bp
            * 23723 24543: contig of 690 bp in length
            * 24543 25379: gap of 100 bp
            * 25379 25479: contig of 715 bp in length
            * 25479 26203: contig of 736 bp in length
            * 26203 26304: gap of 100 bp
            * 26304 27015: contig of 724 bp in length
            * 27015 27115: gap of 100 bp
            * 27115 27845: contig of 712 bp in length
            * 27845 27945: gap of 100 bp
            * 27945 28666: contig of 730 bp in length
            * 28666 28766: gap of 100 bp
            * 28766 29511: contig of 721 bp in length
            * 29511 29612: gap of 100 bp
            * 29612 30339: contig of 745 bp in length
            * 30339 31153: gap of 100 bp
            * 31153 31253: contig of 727 bp in length
            * 31253 31982: gap of 100 bp
            * 31982 31254: contig of 715 bp in length
            * 31254 31982: contig of 729 bp in length
```





```

AE015450_05      500001      610000
AE015450_06      600001      710000
AE015450_07      700001      810000
AE015450_08      800001      910000
AE015450_09      900001      996422
Continuation (5 of 10) of AE015450 Mycoplasma gallisepticum s

Query Match      62.0%; Score:20.6; DB 15; Length 110000;
Best Local Similarity 52.6%; Pred.No.7.4e+02;
Matches 20; Conservative 7; Mismatches 11; Indels 0; Gaps 0;

Qy 5 AYGNTGGCCNTCCNTAARCAATGAYTAYGARGNNAAR 42
Db 88994 ATGCTGACAATCACCTTAATAAGACAATGTAGTAAAA 89031

RESULT 90
AM050691/c
LOCUS
DEFINITION      AM050691      112585 bp      DNA      linear      HTG 01-FEB-2006
PROGRESS ***, 30 unordered pieces.
ACCESSION      AM050691      GI:86439755
VERSION
KEYWORDS      HTG; HTGS PHASE1.
SOURCE      Triticum aestivum (bread wheat)
ORGANISM      Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP
clade; Pooleae; Triticeae; Triticum.

REFERENCE
AUTHORS      Griffiths,S., Sharp,R.R., Foote,T.N., Bertin,I., Wanous,M.,
Reader,S. and Moore,G.
TITLE      Molecular dissection of the Ph1 locus
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 112585)
AUTHORS      Griffiths,S.
TITLE      Direct Submission
JOURNAL      Submitted (19-SEP-2005) Griffiths S., Crop Genetics, John Innes
Centre, Norwich Research Park, Colney, Norwich, NR4 7UH, UNITED
KINGDOM

* NOTE: This is a 'working draft' sequence. It currently
* consists of 30 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1354: contig of 1354 bp in length
* 1355      1454: gap of unknown length
* 1455      2800: contig of 1346 bp in length
* 2801      2900: gap of unknown length
* 2901      4245: contig of 1345 bp in length
* 4246      4345: gap of unknown length
* 4346      5714: contig of 1369 bp in length
* 5715      5815: gap of unknown length
* 5815      7159: contig of 1345 bp in length
* 7160      7259: gap of unknown length
* 7260      8576: contig of 1317 bp in length
* 8577      8676: gap of unknown length
* 8677      10015: contig of 1339 bp in length
* 10016     11626: contig of 1511 bp in length
* 11627     11727: gap of unknown length
* 11727     13952: contig of 2226 bp in length
* 13953     14052: gap of unknown length
* 14053     15566: contig of 1514 bp in length
* 15567     15666: gap of unknown length
* 15667     17035: contig of 1369 bp in length
* 17036     17135: gap of unknown length
* 17136     18227: contig of 1092 bp in length
* 18228     18327: gap of unknown length
* 18328     19508: contig of 1181 bp in length

19509      19608: gap of unknown length
19609      22780: contig of 3172 bp in length
22781      22880: gap of unknown length
22881      25387: contig of 2507 bp in length
25388      25487: gap of unknown length
25488      30548: contig of 4961 bp in length
30549      36148: contig of 5600 bp in length
36149      36248: gap of unknown length
36249      40864: contig of 4616 bp in length
40865      46612: contig of 5648 bp in length
46613      46712: gap of unknown length
46713      50748: contig of 4036 bp in length
50749      57119: contig of 8271 bp in length
57120      57219: gap of unknown length
57220      61870: contig of 4651 bp in length
61871      61970: gap of unknown length
61971      65720: contig of 3750 bp in length
65721      72158: gap of unknown length
72159      72258: gap of unknown length
72259      78415: contig of 6157 bp in length
78416      78515: gap of unknown length
78516      88335: contig of 9820 bp in length
88336      88435: gap of unknown length
88436      97463: contig of 9028 bp in length
97464      97563: gap of unknown length
97564      109677: contig of 12114 bp in length
109678     109777: gap of unknown length
109778     111105: contig of 1328 bp in length
111106     111205: gap of unknown length
111206     112585: contig of 1380 bp in length.

Location/Qualifiers
1. 112585
/organism="Triticum aestivum"
/mol_type="genomic DNA"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/chromosome="5AL"
/clone="606L2"
1355. 1454
/estimated_length=unknown
2801. 2900
/estimated_length=unknown
4246. 4345
/estimated_length=unknown
5715. 5814
/estimated_length=unknown
7160. 7259
/estimated_length=unknown
8577. 8676
/estimated_length=unknown
10016. 10115
/estimated_length=unknown
11627. 11726
/estimated_length=unknown
13953. 14052
/estimated_length=unknown
15567. 15666
/estimated_length=unknown
17036. 17135
/estimated_length=unknown
18228. 18327
/estimated_length=unknown
19509. 19608
/estimated_length=unknown
22781. 22880
/estimated_length=unknown
25388. 25487
/estimated_length=unknown
30449. 30548

```



DeArellano, K., Dewar, K., Domino, M., Doyle, M., Fenestor, J., Ferreira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G., Locke, K., MacDonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., Olivar, T.M., Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D., Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N., Scojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.

# Direct Submission

Submitted (05-FEB-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

## REFERENCE JOURNAL

## AUTHORS

3 (bases 1 to 154471)  
 Birren, B., Linton, L., Nussbaum, C., Lander, E., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Karatas, A., LaRocque, K., Lamaze, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C., MacDonald, P., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P., Sougniez, C., Spencer, B., Stange-Thomann, N., Scojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

## REFERENCE JOURNAL

## AUTHORS

Submitted (01-JUN-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

## REFERENCE JOURNAL

## AUTHORS

4 (bases 1 to 154471)  
 Birren, B., Linton, L., Nussbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gird, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamaze, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Scojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

## REFERENCE JOURNAL

## AUTHORS

Submitted (11-DEC-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

## COMMENT

On Dec 11, 2001 this sequence version replaced gi:14277301.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute / MIT Center for Genome Research

Center code: WIBR  
 Web site: <http://www-seq.wi.mit.edu>  
 Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
 ----- Project Information  
 Center project name: L5905  
 Center clone name: 183\_E24

## FEATURES

source	Location/Qualifiers
1..154471	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="15"
	/map="15"
	/clone="RP11-183E24"
	/clone_lib="RPC1-11 Human Male BAC"
1..281	/rpt_family="MLT2B1"
	complement(302..780)
	/rpt_family="L1MEC"
	complement(787..919)
	/rpt_family="AluSp"
	complement(924..1150)
	/rpt_family="L1MEC"
	complement(1151..1413)
	/rpt_family="AluSg"
	complement(1414..1486)
	/rpt_family="L1MEC"
	1510..1632
	/rpt_family="AT rich"
	complement(1637..1812)
	/rpt_family="L1MEC"
	2014..2155
	/rpt_family="Charlie8"
	3408..4029
	/rpt_family="MLT2F"
	complement(4506..5313)
	/rpt_family="L1MA8"
	complement(5314..5733)
	/rpt_family="WMBT"
	complement(5734..6559)
	/rpt_family="L1MA8"
	complement(6622..6823)
	/rpt_family="MIR"
	10525..10568
	/rpt_family="TG)n"
	11002..11619
	/rpt_family="L1PA14"
	complement(12662..12984)
	/rpt_family="L1MC4a"
	13140..13335
	/rpt_family="L3"
	13354..13482
	/rpt_family="TG)n"
	13484..13563
	/rpt_family="L3"
	complement(14210..14403)
	/rpt_family="MER20"
	complement(14450..14730)
	/rpt_family="L2"
	complement(16431..16801)
	/rpt_family="L1ME3"
	17102..17280
	/rpt_family="MER5A"
	complement(17361..17482)
	/rpt_family="L2"
	17692..17745
	/rpt_family="L2"
	complement(17786..17909)
	/rpt_family="L2"
	complement(18118..18554)
	/rpt_family="MER74C"
	19996..20304

```

/rpt_family="AluSx"
complement(20444..20816)
/rpt_family="MLT11"
complement(20844..20899)
/rpt_family="L3"
complement(20900..21212)
/rpt_family="AluYb8"
complement(21217..21516)
/rpt_family="AluSp"
complement(21517..21651)
/rpt_family="L3"
21653..21846
/rpt_family="MIR"
complement(22812..22893)
/rpt_family="MIR3"
24134..24320
/rpt_family="MER63A"
24956..25064
/rpt_family="MIR"
25728..26030
/rpt_family="AluJo"
26046..26636
/rpt_family="L1MC3"
complement(27077..27379)
/rpt_family="AluSx"
complement(27670..27700)
/rpt_family="L3"
28203..28304
/rpt_family="T-rich"
complement(28334..28523)
/rpt_family="HAL1"
complement(28629..28913)
/rpt_family="HAL1"
complement(29128..29461)
/rpt_family="MLT1A1"

```

```

Query Match      62.0%; Score 20.6; DB 5; Length 154471;
Best Local Similarity 57.1%; Pred. No. 7.9e+02;
Matches 20; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

```

```

QY 4 CAYGNTGGCCNTCCNTAYARGAYTAGARGT 38
Db 135973 CARGTCTGCCATCTCCCTTCAAAGATGCTAAAGT 135939

```

```

RESULT 93
CT025647      165165 bp      DNA      linear      HTG 22-NOV-2005
LOCUS      Danio rerio chromosome 4 clone CH211-173P13, WORKING DRAFT
DEFINITION      SEQUENCE.
ACCESSION      CT025647
VERSION      CT025647.6 GI:82617448
KEYWORDS      HTG; HTGS PHASE2; HTGS ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
      Cypriniformes; Cyprinidae; Danio.
REFERENCE      1 (bases 1 to 165165)
      Matthews,L.
      Direct Submission
      Submitted (21-NOV-2005) Wellcome Trust Sanger Institute, Hinxton,
      Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
      zfish-help@sanger.ac.uk Clone requests:
      http://www.sanger.ac.uk/projects/D.rerio/fags.shtml#dataeight
      On Nov 22, 2005 this sequence version replaced gi:82465356.
      ----- Genome Center
      Center: Wellcome Trust Sanger Institute
      Center code: SC
      Web site: http://www.sanger.ac.uk
      Contact: zfish-help@sanger.ac.uk
      ----- Project Information
      Center project name: zC173P13

```

```

----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 164875 bases at least Q40
Consensus quality: 164972 bases at least Q30
Consensus quality: 165058 bases at least Q20
Insert size: 165165; sum-of-contigs
Insert size: 176473; 13.2% error; agarose-fp
Quality coverage: 9.20x in Q20 bases; sum-of-contigs Quality
coverage: 8.72x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
* 1 165165: contig of 165165 bp in length.
      Location/Qualifiers
          1..165165
            /organism="Danio rerio"
            /mol_type="genomic DNA"
            /db_xref="taxon:7955"
            /chromosome="4"
            /clone="CH211-173P13"
            /clone_lib="CHORI-211"
            1..165165
            /note="assembly_fragment:00095
            clone_end:SP6
            vector_side:left
            clone_end:T7
            vector_side:right"

```

## FEATURES

Source

## misc\_feature

```

1..165165
/note="assembly_fragment:00095
clone_end:SP6
vector_side:left
clone_end:T7
vector_side:right"

```

## ORIGIN

```

Query Match      62.0%; Score 20.6; DB 12; Length 165165;
Best Local Similarity 57.1%; Pred. No. 8e+02;
Matches 20; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

```

```

QY 4 CAYGNTGGCCNTCCNTAYARGAYTAGARGT 38
Db 99162 CATGCAATGCCATCAGCTACAAAGATTACCAAGT 99196

```

## RESULT 94

AP003651/c

```

LOCUS      171116 bp      DNA      linear      HTG 20-DEC-2005
DEFINITION      Mus musculus chromosome 16 clone RP23-19A14 map 16q, *** SEQUENCING
      IN PROGRESS ***, 44 unordered pieces.

```

ACCESSION AP003651

VERSION AP003651.1 GI:83758728

KEYWORDS HTG; HTGS PHASE1.

SOURCE HTG; Mus musculus (house mouse)

## ORGANISM

```

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.

```

## REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

```

1 Hattori,M., Toyoda,A., Taylor,T.D., Fujiyama,A., Yada,T.,
  Totoki,Y., Watanabe,H. and Sakaki,Y.
  Mus musculus genomic DNA
  Published Only in Database (2005)
  2 (bases 1 to 171116)
  Hattori,M., Toyoda,A., Taylor,T.D., Fujiyama,A., Yada,T.,
  Totoki,Y., Watanabe,H. and Sakaki,Y.
  Direct Submission
  Submitted (22-MAY-2001) Masahira Hattori, The Institute of Physical
  and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
  1-7-22 Suehiro-chou, Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
  (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
  Tel:81-45-503-9111, Fax:81-45-503-9170)

```



## COMMENT

NOTE: This is a 'working draft' sequence. It currently consists of 44 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

```
1      7288 contig of 7288 bp in length
      7389
      15472
      25490 contig of 7983 bp in length
      25591
      33522 contig of 7932 bp in length
      33623
      39902 contig of 6280 bp in length
      40003
      47009 contig of 7007 bp in length
      47110
      55409 contig of 8300 bp in length
      55510
      62347 contig of 6838 bp in length
      62448
      68266
      74618
      80349 contig of 5732 bp in length
      80449
      80350
      85274
      85275
      85374
      85375
      90541
      90542
      90641
      90642
      95715
      95716
      95815
      95816
      95978
      99979
      100079
      105151
      105152
      105251
      109307
      109308
      109309
      109407
      109408
      112997
      112998
      113097
      113098
      116822
      116823
      116923
      116924
      121122
      121123
      121223
      124849
      124850
      124949
      124950
      129645
      129646
      129746
      131801
      131802
      131901
      131902
      134233
      134234
      136774
      136775
      139997
      139998
      140097
      140098
      142968
      142969
      143069
      143070
      145514
      145515
      145615
      147824
      147825
      147924
      150151
      150152
      150252
      151925
      151926
      152025
      152026
      153636
      153637
      153736
      155687
      155688
      155788
      157365
      157366
      157466
      159194
      159195
      159294
      160838
      160839
      160939
      162269
      162270
      162370
      163441
      163442
      163542
      165038
      165138
      166363
      166463
      167581
      167582
      168768
      168769
      169871
      169872
      169972
      171116
      171117
      7288: contig of 7288 bp in length
      7389: gap of 100 bp
      15371: contig of 7983 bp in length
      15372
      15471: gap of 100 bp
      25490: contig of 10019 bp in length
      25491
      33522: contig of 7932 bp in length
      33523
      33623: gap of 100 bp
      33624
      39902: contig of 6280 bp in length
      39903
      40002: gap of 100 bp
      40003
      47009: contig of 7007 bp in length
      47109: gap of 100 bp
      55409: contig of 8300 bp in length
      55410
```

\* NOTE: This is a 'working draft' sequence. It currently consists of 44 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```
* 55510
* 62347: contig of 6838 bp in length
* 62447: gap of 100 bp
* 68165: contig of 5718 bp in length
* 68265: gap of 100 bp
* 74617: contig of 6252 bp in length
* 74618: gap of 100 bp
* 80349: contig of 5732 bp in length
* 80449: gap of 100 bp
* 85274: contig of 4825 bp in length
* 85374: gap of 100 bp
* 85375
* 90541: contig of 5167 bp in length
* 90641: gap of 100 bp
* 95715: contig of 5074 bp in length
* 95815: gap of 100 bp
* 95978: contig of 4163 bp in length
* 99979: gap of 100 bp
* 100079
* 105151: contig of 5073 bp in length
* 105251: gap of 100 bp
* 109307: contig of 4056 bp in length
* 109308: gap of 100 bp
* 109407: gap of 100 bp
* 109408
* 112997: contig of 3590 bp in length
* 112998: gap of 100 bp
* 113097: gap of 100 bp
* 113098
* 116822: contig of 3725 bp in length
* 116923: gap of 100 bp
* 121122: contig of 4200 bp in length
* 121223: gap of 100 bp
* 124849: contig of 3627 bp in length
* 124949: gap of 100 bp
* 124950
* 129645: contig of 4696 bp in length
* 129646: gap of 100 bp
* 129746
* 131801: contig of 2056 bp in length
* 131901: gap of 100 bp
* 131902
* 134233: contig of 2332 bp in length
* 134334: gap of 100 bp
* 136774: contig of 2440 bp in length
* 136775: gap of 100 bp
* 139997: contig of 3124 bp in length
* 139998: gap of 100 bp
* 140097: gap of 100 bp
* 140098
* 142968: contig of 2871 bp in length
* 143068: gap of 100 bp
* 143069
* 145514: contig of 2446 bp in length
* 145515: gap of 100 bp
* 145615
* 147823: contig of 2209 bp in length
* 147824
* 147923: gap of 100 bp
* 150151: contig of 2228 bp in length
* 150251: gap of 100 bp
* 150252
* 151924: contig of 1673 bp in length
* 151925
* 152024: gap of 100 bp
* 152025
* 153635: contig of 1611 bp in length
* 153636: gap of 100 bp
* 153736
* 155687: contig of 1952 bp in length
* 155688
* 155788: gap of 100 bp
* 157365: contig of 1578 bp in length
* 157366: gap of 100 bp
* 157466
* 159194: contig of 1729 bp in length
* 159195
* 159294: gap of 100 bp
* 159295
* 160838: contig of 1544 bp in length
* 160839
* 160939: contig of 1331 bp in length
* 162269: contig of 1331 bp in length
* 162270
* 162370: gap of 100 bp
* 163441: contig of 1072 bp in length
* 163442
* 163541: gap of 100 bp
* 163542
* 165037: contig of 1496 bp in length
* 165137: gap of 100 bp
* 165138
* 166362: contig of 1225 bp in length
* 166363
* 166462: gap of 100 bp
* 166463
* 167580: contig of 1118 bp in length
* 167581
* 167680: gap of 100 bp
* 167681
* 168767: contig of 1087 bp in length
* 168768
* 168867: gap of 100 bp
* 168868
* 169871: contig of 1004 bp in length
* 169872
* 169971: gap of 100 bp
* 171116: contig of 1145 bp in length.
* 169972
```



----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC

```
repeat_region /rpt_family="MaLR"
14884..16083
repeat_region /rpt_family="L1"
16092..16331
repeat_region /rpt_family="L1"
16333..16575
repeat_region /rpt_family="MaLR"
1756..19175
repeat_region /rpt_family="L1"
23186..23271
repeat_region /rpt_family="L1"
23375..23765
repeat_region /rpt_family="L1"
23766..24913
repeat_region /rpt_family="ERVK"
24926..25072
repeat_region /rpt_family="ERVK"
25089..25393
repeat_region /rpt_family="ERVK"
25824..26056
repeat_region /rpt_family="L1"
26104..27070
repeat_region /rpt_family="L1"
27063..27699
repeat_region /rpt_family="L1"
27703..27935
repeat_region /rpt_family="MaLR"
27936..28264
repeat_region /rpt_family="MaLR"
28282..28592
repeat_region /rpt_family="L1"
29879..31448
repeat_region /rpt_family="MaLR"
31460..31969
repeat_region /rpt_family="L1"
32044..32177
repeat_region /rpt_family="Alu"
32100..32194
repeat_region /rpt_family="B4"
32212..32263
repeat_region /rpt_family="L1"
33702..33767
repeat_region /rpt_family="B4"
34184..34276
repeat_region /rpt_family="Alu"
35093..35241
repeat_region /rpt_family="B4"
35658..35846
repeat_region /rpt_family="L1"
35835..36421
repeat_region /rpt_family="L1"
36750..36962
repeat_region /rpt_family="MaLR"
37098..40719
repeat_region /rpt_family="L1"
40713..42128
repeat_region /rpt_family="L1"
42729..42865
repeat_region /rpt_family="MaLR"
43147..43333
repeat_region /rpt_family="B2"
43432..43581
repeat_region /rpt_family="L1"
44196..44397
repeat_region /rpt_family="B4"
44652..44842
repeat_region /rpt_family="ERVK"
44757..44990
repeat_region /rpt_family="B4"
45301..45394
repeat_region /rpt_family="L1"
45466..45894
repeat_region /rpt_family="L1"

repeat_region 47361..47698
/rpt_family="ERVK"
54091..54444
/rpt_family="ERVK"
54988..55095
/rpt_family="MER1_type"
57030..57250
/rpt_family="L1"
57387..57797
/rpt_family="L1"
57837..58025
/rpt_family="B2"

Query Match 62.0%; Score 20.6; DB 6; Length 191693;
Best Local Similarity 64.5%; Pred. No. 8.2e+02;
Matches 20; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 8 CNTGGCNCNCNTAYAAAGAYTAYGARGT 38
||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 172540 CCTGGCATTCCCATATAAAGAAATATAAGT 172570

RESULT 98
AC147235/c 193668 bp DNA linear ROD 27-JAN-2005
LOCUS Mus musculus BAC clone RP23-353B7 from 6, complete sequence.
DEFINITION AC147235
ACCESSION AC147235.3 GI:54292044
VERSION HTG.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus

REFERENCE
1 (bases 1 to 193668)
AUTHORS Zheng,X., Cotton,M., Haglund,K. and Meyer,R.
TITLE The sequence of Mus musculus BAC clone RP23-353B7
JOURNAL Unpublished (2001)
REFERENCE 2 (bases 1 to 193668)
AUTHORS Wilson,R.K.
TITLE Direct Submission
JOURNAL Submitted (07-NOV-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 3 (bases 1 to 193668)
AUTHORS Wilson,R.K.
TITLE Direct Submission
JOURNAL Submitted (17-AUG-2004) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 193668)
AUTHORS Wilson,R.K.
TITLE Direct Submission
JOURNAL Submitted (19-OCT-2004) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 193668)
AUTHORS Wilson,R.K.
TITLE Direct Submission
JOURNAL Submitted (27-JAN-2005) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Oct 19, 2004 this sequence version replaced gi:51315670.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@watson.wustl.edu
----- Summary Statistics
-----
Center project name: M_BA0353B07

NOTICE:
This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
```

chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

#### SOURCE INFORMATION:

The RPCI-23 BAC Library has been constructed by Kazutoyo Oseawa and Minako Tatenio in the laboratory of Pieter de Jong (<http://www.chori.org>) from female C57BL/6J mouse kidney and/or brain genomic DNA. The clone and detailed information can be obtained from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

#### NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone.

#### Location/Qualifiers

```
1. 193668
   /organism="Mus musculus"
   /mol_type="genomic DNA"
   /db_xref="taxon:10090"
   /chromosome="6"
   /map="6"
   /clone="RP23-353B7"
   /clone_lib="RPCI-23"
   149359..149522
   /note="Sequence derived from PCR product of project DNA."
```

#### FEATURES

source

#### ORIGIN

```
Query Match      62.0%; Score 20.6; DB 6; Length 193668;
Best Local Similarity 64.5%; Pred. No. 8.2e+02;
Matches 20; Conservative 4; Mismatches 7; Indels 0; Gaps 0;
```

QY 8 CNTGGCCNTCCNTAAGCATATGARGT 38

Db 176270 CTTGGCATCCCCATATAAGATATAAAGT 176240

#### RESULT 99

AC127132

DEFINITION Rattus norvegicus clone CH230-156A5, WORKING DRAFT SEQUENCE, 2  
unordered pieces.

AC127132.3 GI:30521218

HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.

Rattus norvegicus (Norway rat)

SOURCE

ORGANISM

Eukaryota; Metazoa; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muroidae; Muridae; Murinae; Rattus.

1 (bases 1 to 207858)

Muzny,D,Marie., Metzker,M, Lee., Abramson,S., Adams,C., Alder,J.,

Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,

Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,

Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,

Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,

Bryant,N., Buay,C., Burch,P., Burrell,K., Calderon,E.,

Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,

Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,

Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,

Davila,M., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,

Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,

Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,

Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,

Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,

Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,

Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,  
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,  
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,  
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hoques,M.,  
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,  
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,  
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,  
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,  
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,  
Lorensuewa,L., Loulsegged,H., Lozado,R.J., Lu,X., Ma,J.,  
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,  
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,  
Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,  
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,  
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,  
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,  
Nwaokemele,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,  
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,  
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L., L.,  
Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,  
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,  
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,  
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,  
Shetty,J., Shvartabeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,  
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,  
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,  
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,  
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,  
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,  
Williams,G., Willson,R., Wlsczyk,R., Wooden,H., Worley,K.,  
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,  
Yu,P., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von  
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,  
Weinstock,G. and Gibbs,R.A.

#### Direct Submission

Unpublished

2 (bases 1 to 207858)

Worley,K.C.

#### Direct Submission

Submitted (13-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 207858)

Rat Genome Sequencing Consortium.

#### Direct Submission

Submitted (10-MAY-2003) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

On May 10, 2003 this sequence version replaced gi:23664503.

The sequence in this assembly is a combination of BAC based reads  
and whole genome shotgun sequencing reads assembled using Atlas  
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described  
in the feature table below represents a scaffold in the Atlas  
assembly (a 'contig-scaffold'). Within each contig-scaffold,  
individual sequence contigs are ordered and oriented, and separated  
by sized gaps filled with Ns to the estimated size. The sequence  
may extend beyond the ends of the clone and there may be sequence  
contigs within a contig-scaffold that consist entirely of whole  
genome shotgun sequence reads. Both end sequences and whole genome  
shotgun sequence only contigs will be indicated in the feature  
table.

#### ----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

----- Project Information

Center project name: G10F

Center clone name: CH230-156A5

----- Summary Statistics

Assembly program: Atlas 3.0;

Consensus quality: 199732 bases at least Q40

Consensus quality: 201456 bases at least Q30



Estimated insert size: 212927; sum-of-contigs estimation  
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 1 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
1 219121: contig of 219121 bp in length.

FEATURES

source

1 .219121

/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-318018"

misc\_feature

1..1616  
/note="wgs\_end\_extension  
clone\_end:T7"

misc\_feature

3372..4247  
/note="clone\_boundary  
clone\_end:T7"  
site:

misc\_feature

end\_sequence:BZ124697"  
complement(215586..216436)  
/note="clone\_boundary  
clone\_end:Sp5  
site:  
end\_sequence:BZ124724"

ORIGIN

Query Match 62.0%; Score 20.6; DB 12; Length 219121;  
Best Local Similarity 56.1%; Pred. No. 8.4e+02;  
Matches 23; Conservative 4; Mismatches 14; Indels 0; Gaps 0;  
OY 1 GCNCAYGCNTGCCNTCCNTAYAAAGAYTAYGAGTNA 41  
|||:|||||:|||||:|||||:|||||:|||||:|||||:  
Db 205673 GCACAGCCTGGCCCTCACCCCATGGAGAGTGAGTGTACA 205633

Search completed: June 10, 2006, 18:19:48  
Job time : 1232.2 secs

**THIS PAGE IS BLANK**



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2006, 15:16:40 ; Search time 312.2 Seconds  
(without alignments)  
937.971 Million cell updates/sec

Title: US-10-600-816-33

Perfect score: 33.2

Sequence: 1 gcnaycgntggcncntccc.....ayaargaytaygargtnaar 42

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 5244920 seqs, 3486124231 residues

Total number of hits satisfying chosen parameters: 10489840

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 1000 summaries

Database :

N Geneseq\_8.\*

1: Geneseqn1980s.\*

2: Geneseqn1990s.\*

3: Geneseqn2000s.\*

4: Geneseqn2001as.\*

5: Geneseqn2001bs.\*

6: Geneseqn2002as.\*

7: Geneseqn2002bs.\*

8: Geneseqn2003as.\*

9: Geneseqn2003bs.\*

10: Geneseqn2003cs.\*

11: Geneseqn2003ds.\*

12: Geneseqn2004as.\*

13: Geneseqn2004bs.\*

14: Geneseqn2005s.\*

15: Geneseqn2006s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	33.2	100.0	42	12	AD128530 Human GPC
2	30	90.4	497	11	ADT96418 Colon can
3	30	90.4	497	11	ADT96418 Colon can
4	30	90.4	552	11	ADT96292 Colon can
5	30	90.4	552	11	ADT96292 Colon can
6	30	90.4	603	3	AAZ90046 Human cDN
7	30	90.4	620	6	ABQ58527 Human col
8	30	90.4	634	6	ABQ58527 Human col
9	30	90.4	642	10	ABT22328 Breast ca
10	30	90.4	1074	12	ADO30035 Human GPC
11	30	90.4	1212	3	AAZ61776 cDNA enco
12	30	90.4	1212	6	AAZ61776 cDNA enco
13	30	90.4	1212	6	AAZ61776 cDNA enco
14	30	90.4	1228	8	ACA04775 Human cDN
15	30	90.4	1460	14	AEA00112 Human TAT
16	30	90.4	1460	14	AEA00112 Human TAT
17	30	90.4	1619	4	AAZ58615 Human REC
18	30	90.4	1718	3	AAZ90056 Hydrophob

19	30	90.4	1788	10	ADF70573	Orphan re
20	30	90.4	2302	6	ABT10173	Human bre
21	30	90.4	2302	8	ACC58386	Human GPC
22	30	90.4	2302	10	ADD93240	RAIG1 cod
23	30	90.4	2302	10	ADL24773	Intestina
24	30	90.4	2316	10	ABT31923	Human bre
25	30	90.4	2446	4	AAH14688	Human cDN
26	30	90.4	2456	8	ABZ42832	Human G p
27	30	90.4	2456	10	ACC72695	Human can
28	30	90.4	2456	11	ADN39301	Cancer/an
29	30	90.4	2456	12	AD128525	Human GPC
30	30	90.4	2456	12	AD128525	Human GPC
31	30	90.4	2456	13	ADR48221	Human ret
32	30	90.4	2456	13	ACN39388	Tumour-as
33	30	90.4	2456	13	ADR43992	Human bre
34	30	90.4	2456	13	ADU06126	Novel bro
35	30	90.4	2456	14	AEA00088	Human TAT
36	30	90.4	2456	14	AEA00088	Human TAT
37	30	90.4	2456	14	AEQ47480	Retinoid
38	30	90.4	2593	6	ABQ54954	Human ova
39	30	90.4	3371	4	AAH72766	Human cer
40	30	90.4	4239	11	ACN89274	Breast ca
41	30	90.4	6730	13	ADN92749	Pancrati
42	27.4	82.5	680	4	AAH08354	Human CDN
43	21.4	64.5	167932	10	ADL13501	Osteoarth
44	21.2	63.9	7880	13	ADR84444	Aspergill
45	20.8	62.7	421	13	ADQ56529	Novel can
46	20.8	62.7	349980	6	ABQ81845	Bifidobac
47	20.6	62.0	2004	2	AAQ51538	Mycobacte
48	20.6	62.0	2577	2	AAQ68672	Partial P
49	20.4	61.4	1053	14	AED71703	Corynebac
50	20.4	61.4	1074	5	AAH65560	C glutam
51	20.4	61.4	1201	11	AD165724	C. glutam
52	20.4	61.4	2780	6	ABQ76459	S. cerevi
53	20.4	61.4	349980	5	AAH68525	C. glutami
54	20.2	60.8	1146	8	ACA37503	Prokaryot
55	20.2	60.8	110000	13	ABD32911	Continuation (3 of
56	20.2	60.8	110000	14	ABE39175	Continuation (23 o
57	20.2	60.8	110000	14	ABE42401	Continuation (23 o
58	20.2	60.8	110000	14	ABE42737	Continuation (9 of
59	20.2	60.8	189430	14	ABE35718	Aeb35718 L. pneumo
60	20.2	60.8	263853	14	ABE39171	Aeb39171 L. pneumo
61	20	60.2	669	4	ABL91353	Chlamydia
62	20	60.2	5161	4	AAZ36060	Human car
63	20	60.2	5161	10	ADZ46754	Human car
64	20	60.2	5161	13	ADJ08172	Human car
65	20	60.2	110000	2	AAZ91990	Continuation (9 of
66	20	60.2	110000	13	ABD32968	Continuation (7 of
67	19.6	59.0	356	14	ADY98830	T. reesei
68	19.6	59.0	691	6	ABQ55988	Human ful
69	19.6	59.0	1083	13	ADX63042	Plant ful
70	19.6	59.0	1305	13	ADX36401	Plant ful
71	19.4	58.4	286	10	ABZ90108	Human nuc
72	19.4	58.4	286	11	ABD26338	AA459692
73	19.4	58.4	354	4	AAK58278	Human imm
74	19.4	58.4	614	10	ACD92866	Human col
75	19.4	58.4	918	6	ABZ57439	Formamido
76	19.4	58.4	918	13	ADW28597	cDNA enco
77	19.4	58.4	992	13	ADT04975	Haemophil
78	19.4	58.4	1757	4	AAH99671	Human pro
79	19.4	58.4	2321	5	ABX71365	Human sig
80	19.4	58.4	2416	6	ABZ11510	Human pol
81	19.4	58.4	2416	12	ADM44028	Novel hum
82	19.4	58.4	3688	3	AAZ47167	Arabidops
83	19.4	58.4	3738	10	ADF01135	Bacterial
84	19.2	57.8	418	2	AAZ11190	Interleuk
85	19.2	57.8	418	6	AAZ39735	Human int
86	19.2	57.8	548	6	AAZ39735	Human int
87	19.2	57.8	550	2	AAZ11191	Interleuk
88	19.2	57.8	1001	3	AAZ37760	Arachidon
89	19.2	57.8	1068	12	ADQ030325	Mouse GPC
90	19.2	57.8	1704	12	ADQ08695	Ciona int
91	19.2	57.8	1828	4	AAZ56974	C. tracho

92	19.2	57.8	1828	10	ADD42768	165	18.6	56.0	1284	8	ACA39199	AcA39199 Prokaryot
93	19.2	57.8	2090	4	ABLI12999	166	18.6	56.0	1305	12	ADL03247	AdL03247 DNA encod
c 94	19.2	57.8	2496	12	ADL12908	c 167	18.6	56.0	1351	14	ADZ60298	AdZ60298 Murine 25
c 95	19.2	57.8	2648	4	AAS34914	c 168	18.6	56.0	1616	13	ADX65542	AdX65542 Plant ful
c 96	19.2	57.8	2648	10	ADC46072	c 169	18.6	56.0	1984	13	ADX62925	AdX62925 Plant ful
c 97	19.2	57.8	3096	8	ACA30557	c 170	18.6	56.0	67167	14	ADY80053	AdY80053 Nucleotid
c 98	19.2	57.8	3397	13	ADV35091	c 171	18.6	56.0	94750	4	AAR28551	AAR28551 Genomic f
99	19.2	57.8	3397	14	ADZ60766	c 172	18.6	56.0	110000	13	ABD32721_1	Continuation (2 of
100	19.2	57.8	3397	14	ADZ60766	c 173	18.6	56.0	130480	4	AAR25833	AAR25833 R. marinu
101	19.2	57.8	3477	10	ABT41712	c 174	18.4	55.4	214	5	ABA11205	ABa11205 Human ner
102	19.2	57.8	4090	4	ABLI12998	c 175	18.4	55.4	320	13	ADR65193	AdR65193 Cotton cD
103	19.2	57.8	12222	4	AAR54045	c 176	18.4	55.4	425	5	ADL41653	AdL41653 Human ova
c 104	19.2	57.8	68193	12	ADQ97400	c 177	18.4	55.4	483	9	ACH39801	ACH39801 Human foe
c 105	19.2	57.8	82660	11	ACN45192	c 178	18.4	55.4	483	13	ADR65192	AdR65192 Cotton cD
c 106	19.2	57.8	83493	14	ADZ13310	c 179	18.4	55.4	519	4	AAH07459	AAH07459 Human cDN
c 107	19.2	57.8	83943	13	ABD32881	c 180	18.4	55.4	588	13	ACN60374	ACn60374 Cotton gy
c 108	19.2	57.8	110000	2	AZ01425	c 181	18.4	55.4	598	13	ADR65190	AdR65190 Cotton cD
c 109	19.2	57.8	110000	13	ABD32968_2	c 182	18.4	55.4	761	10	ADL06700	AdL06700 Carassius
c 110	19.2	57.8	156321	14	ACG82893	c 183	18.4	55.4	887	5	AAS66149	AAS66149 DNA encod
c 111	19.2	57.8	173805	10	ADL13775	c 184	18.4	55.4	910	12	ADL14030	AdL14030 Human cDN
c 112	19.2	57.8	322101	10	ADZ58431	c 185	18.4	55.4	1128	14	ABD24357	ABD24357 Sand rat
c 113	19.2	57.8	322101	12	AD135046	c 186	18.4	55.4	1600	3	AAC47340	AAc47340 Arabidops
c 114	19	57.2	184	3	AAAC17478	c 187	18.4	55.4	1604	3	AAC40491	AAc40491 Arabidops
c 115	19	57.2	2232	8	ACA29136	c 188	18.4	55.4	1644	13	ADR24452	AdR24452 Breast ca
c 116	19	57.2	3165	8	ABX13154	c 189	18.4	55.4	1663	12	ADO35460	ADO35460 Novel mou
c 117	19	57.2	3915	5	AAH77995	c 190	18.4	55.4	2325	6	ABQ69466	ABQ69466 Listeria
c 118	19	57.2	4112	5	ABV24426	c 191	18.4	55.4	2376	6	ABQ67859	ABQ67859 Listeria
c 119	19	57.2	4136	6	ABK10102	c 192	18.4	55.4	2397	10	AAAD51688	AAAD51688 Human nuc
c 120	19	57.2	4151	5	ABV24466	c 193	18.4	55.4	2485	13	ADX11038	AdX11038 Plant ful
c 121	19	57.2	109565	13	ABD33086	c 194	18.4	55.4	2520	4	AAH34235	AAH34235 Human col
c 122	19	57.2	110000	2	AZ01425_01	c 195	18.4	55.4	2522	6	ABX71777	ABx71777 Human dit
c 123	18.8	56.6	255	4	AAK57623	c 196	18.4	55.4	2526	4	AAI59777	AAi59777 Human pol
c 124	18.8	56.6	333	4	AAK530881	c 197	18.4	55.4	2830	4	AAI57991	AAi57991 Human pol
c 125	18.8	56.6	349	4	AAI84360	c 198	18.4	55.4	2843	12	ADQ84108	ADQ84108 Human tnm
c 126	18.8	56.6	413	10	ADD26877	c 199	18.4	55.4	2845	4	AAH18012	AAH18012 Human cDN
c 127	18.8	56.6	594	6	ABNG1380	c 200	18.4	55.4	2963	13	ACN42579	ACn42579 Human dia
c 128	18.8	56.6	948	6	ABNG1380	c 201	18.4	55.4	2967	13	ACN42578	ACn42578 Human dia
c 129	18.8	56.6	2000	11	ACL35075	c 202	18.4	55.4	3151	9	ADB17497	ADB17497 Wheat pos
c 130	18.8	56.6	2000	11	ACL37468	c 203	18.4	55.4	3151	14	ABC75721	ABC75721 Wheat Ar9
c 131	18.8	56.6	2933	12	ADO36051	c 204	18.4	55.4	5868	4	AAK89398	AAK89398 Human dig
c 132	18.8	56.6	3396	2	AAO55515	c 205	18.4	55.4	7930	4	ABL10650	ABl10650 Drosophil
c 133	18.8	56.6	5603	4	ABLI12539	c 206	18.4	55.4	10316	4	AAK89399	AAK89399 Human dig
c 134	18.8	56.6	10327	4	ABLI12538	c 207	18.4	55.4	18981	4	ABL04961	ABl04961 Human rep
c 135	18.8	56.6	11566	5	ABAI4380	c 208	18.4	55.4	18981	4	ABL97854	ABl97854 Human tes
c 136	18.8	56.6	12000	6	ABK34015	c 209	18.4	55.4	22651	4	AAK78202	AAK78202 Human imm
c 137	18.8	56.6	12963	8	ACF62802	c 210	18.4	55.4	31337	12	ADQ97031	ADQ97031 Mouse can
c 138	18.8	56.6	12963	8	ACF62802	c 211	18.4	55.4	81001	4	AAF30035	AAF30035 Human apo
c 139	18.8	56.6	12963	13	ADS89640	c 212	18.4	55.4	110000	6	ABO69245_10	Continuation (11 o
c 140	18.8	56.6	12963	13	ADS89366	c 213	18.4	55.4	110000	6	ABO67197_09	Continuation (10 o
c 141	18.8	56.6	14362	6	ABQ67048	c 214	18.4	55.4	110000	11	ACN44014_1	Continuation (2 of
c 142	18.8	56.6	14708	4	AA545514	c 215	18.4	55.4	110000	14	AEA61120_3	Continuation (4 of
c 143	18.8	56.6	14708	6	ABL34245	c 216	18.4	55.4	110000	14	AEA61102_1	Continuation (2 of
c 144	18.8	56.6	14708	6	ABL923325	c 217	18.4	55.4	161671	12	ADH30232	AdH30232 Human PLO
c 145	18.8	56.6	14708	6	ABK28450	c 218	18.4	55.4	174318	15	AEF92655	AEf92655 Human Ras
c 146	18.8	56.6	15863	10	ADB54294	c 219	18.4	55.4	188971	12	ADL08108	AdL08108 Human gen
c 147	18.8	56.6	15863	10	ADB54166	c 220	18.4	55.4	200000	12	ADO47193	ADO47193 DNA seque
c 148	18.8	56.6	23483	13	ABD32835	c 221	18.4	55.4	202251	11	ACN44504	ACn44504 Mouse gen
c 149	18.8	56.6	106323	8	ABX14591	c 222	18.4	55.4	277616	13	ABD32602	ABD32602 Human can
c 150	18.8	56.6	106323	10	ADN458821	c 223	18.2	54.8	171	2	AAK83259	AAK83259 Breast ca
c 151	18.8	56.6	110000	12	ADN46845_09	c 224	18.2	54.8	171	2	AAV68857	AAV68857 DNA molec
c 152	18.8	56.6	110000	12	ADN47591_11	c 225	18.2	54.8	171	3	AAK80782	AAK80782 Human bre
c 153	18.8	56.6	110000	12	ADN46123_09	c 226	18.2	54.8	171	6	AAK99628	AAK99628 Breast tu
c 154	18.8	56.6	110000	12	ADN47209_11	c 227	18.2	54.8	171	6	ABK46672	ABk46672 Human bre
c 155	18.8	56.6	110000	12	ADN46464_09	c 228	18.2	54.8	171	8	ADAI1139	ADaI1139 Human bre
c 156	18.8	56.6	110000	12	ADN47960_11	c 229	18.2	54.8	171	10	ADC15112	ADC15112 Human bre
c 157	18.6	56.0	23	12	ADI28532	c 230	18.2	54.8	176	4	AAI25240	AAi25240 Probe #15
c 158	18.6	56.0	265	6	ABN26856	c 231	18.2	54.8	176	4	ABA71057	ABa71057 Human foe
c 159	18.6	56.0	400	2	ABW77236	c 232	18.2	54.8	176	4	AAI51271	AAi51271 Probe #19
c 160	18.6	56.0	438	8	ABZ53830	c 233	18.2	54.8	176	4	ABA37448	ABa37448 Probe #15
c 161	18.6	56.0	725	13	ADR13190	c 234	18.2	54.8	176	4	AAK45317	AAK45317 Human bon
c 162	18.6	56.0	906	10	ADP14188	c 235	18.2	54.8	176	4	AAK19339	AAK19339 Human bra
c 163	18.6	56.0	934	13	ADX60641	c 236	18.2	54.8	176	4	ABS44996	ABs44996 Human liv
c 164	18.6	56.0	1240	13	ADX11208	c 237	18.2	54.8	176	6	ABS19574	ABs19574 Human gen

238	18.2	54.8	395	9	ACH49830	Ach49830 Human leu	311	18	54.2	878	2	AAV16672	AAV16672 Polynucle
239	18.2	54.8	410	6	AAL13639	Aal13639 Human bre	312	18	54.2	878	6	ABK15460	Abk15460 Human col
240	18.2	54.8	449	6	ABN64231	Abn64231 Human can	313	18	54.2	878	6	ABK52542	Abk52542 Human col
241	18.2	54.8	450	11	ACN83709	Acn83709 Breast ca	c 314	18	54.2	1000	13	ADX51155	Adx51155 Plant ful
242	18.2	54.8	470	4	AAL16024	Aal16024 Probe #59	315	18	54.2	1000	15	AEF24500	Aef24500 Horse mit
243	18.2	54.8	470	4	ABA58464	Aba58464 Human foe	316	18	54.2	1000	15	AEF24497	Aef24497 Horse mit
244	18.2	54.8	470	4	AAL138116	Aal138116 Probe #68	317	18	54.2	1000	15	AEF24496	Aef24496 Horse mit
245	18.2	54.8	470	4	ABA27543	Aba27543 Probe #60	318	18	54.2	1029	8	ADA69628	Ada69628 Rice gene
246	18.2	54.8	470	4	AAK33263	Aak33263 Human bon	319	18	54.2	1029	11	ACL27148	Acl27148 Rice abio
247	18.2	54.8	470	4	AAK06574	Aak06574 Human bra	320	18	54.2	1080	8	ACA51895	Aca51895 Prokaryot
248	18.2	54.8	470	4	ABE311967	Abes311967 Human liv	321	18	54.2	1080	8	ACA19692	Acac19692 Prokaryot
249	18.2	54.8	470	6	ABSO7038	Abso7038 Human gen	322	18	54.2	1081	8	ACA48632	Acac48632 Prokaryot
250	18.2	54.8	623	12	ADJ43982	Adj43982 Plant cDN	323	18	54.2	1165	5	AAH76179	Aah76179 DNA encod
251	18.2	54.8	669	14	ACL65099	Acl65099 M. xanthu	324	18	54.2	1165	5	AAH76150	Aah76150 DNA encod
252	18.2	54.8	864	9	ADA29982	Ada29982 DNA encod	325	18	54.2	1191	8	ABZ56934	Abz56934 Human RGS
253	18.2	54.8	1101	9	ADA330049	Ada330049 DNA encod	326	18	54.2	1305	13	ADR60022	Adr60022 Cotton CD
254	18.2	54.8	1284	8	ACA33044	Acac33044 Prokaryot	c 327	18	54.2	1341	8	ACA30486	Acac30486 Prokaryot
255	18.2	54.8	1335	10	ADH84124	Adh84124 Enterococ	c 328	18	54.2	1467	13	ADT47812	Adt47812 Bacterial
256	18.2	54.8	1594	10	ADJ56323	Adj56323 Human cDN	329	18	54.2	1512	10	ACA56252	Acac56252 Cow signa
257	18.2	54.8	2000	10	ACC60653	Acc60653 Gene sequ	330	18	54.2	1512	12	ADI56048	Adi56048 Human pol
258	18.2	54.8	2000	10	ADK64153	Adk64153 Disease t	331	18	54.2	1632	5	AAS68142	Aas68142 DNA encod
259	18.2	54.8	2115	13	ADS46677	Ads46677 Bacterial	332	18	54.2	1662	4	AAS06702	Aas06702 Polynucle
260	18.2	54.8	4365	6	AAS94791	Aas94791 Human DNA	333	18	54.2	1662	4	AAH78797	Aah78797 Human G-p
261	18.2	54.8	4416	3	AACS9319	Aacs9319 Human sec	334	18	54.2	1662	6	ABQ86178	Abq86178 Novel hum
262	18.2	54.8	4417	3	ABL01582	Ablo1582 Human sec	335	18	54.2	1662	6	AAD26466	Aad26466 Human kin
263	18.2	54.8	29729	2	AAI13175	Aai13175 Enterococ	336	18	54.2	1662	6	AAD28071	Aad28071 Human kin
264	18.2	54.8	29729	6	ABS98970	Abs98970 Enterococ	337	18	54.2	1662	6	AAD46350	Aad46350 Human con
265	18.2	54.8	64829	14	AEA61181	Aea61181 Human FHL	338	18	54.2	1662	8	ACC44826	Acc44826 Human G p
266	18.2	54.8	80963	14	AEA61095	Aea61095 Continuation (4 of	339	18	54.2	1662	12	ADG12799	Adg12799 Human GRK
267	18.2	54.8	110000	13	ABD32629	Abd32629 Continuation (4 of	340	18	54.2	1662	13	ADG60136	Adg60136 Human G p
268	18.2	54.8	110000	14	AEA42401	Aea42401 Continuation (29 o	341	18	54.2	1683	15	AEH28185	Aeh28185 Human CS1
269	18.2	54.8	112414	6	ABL59091	Ably59091 Nucleotid	342	18	54.2	1683	15	AEH28220	Aeh28220 Human CS1
270	18.2	54.8	127722	12	ADQ97301	Adq97301 Human can	343	18	54.2	1701	6	ADI16737	Adi16737 Human NOV
271	18.2	54.8	152811	14	AEH18413	Aeh18413 Fibrotic	344	18	54.2	1761	8	ABZ56943	Abz56943 Human gen
272	18.2	54.8	160671	14	AEH04683	Aeh04683 Cancer-as	345	18	54.2	1821	10	ADC39221	Adc39221 Novel hum
273	18	54.2	230	11	ADT96102	Adt96102 Colon can	346	18	54.2	1950	10	ABX07405	Abx07405 S. pneumo
274	18	54.2	230	11	ADX42584	Adx42584 Human cDN	347	18	54.2	1953	8	ABZ42476	Abz42476 Streptoco
275	18	54.2	242	15	AEH28182	Aeh28182 Human CS1	348	18	54.2	1953	13	ADR92277	Adr92277 Novel S.
276	18	54.2	255	2	AAT22483	Aat22483 Human gen	349	18	54.2	1953	13	ADK44045	Adk44045 Streptoco
277	18	54.2	270	5	AAS75287	Aas75287 DNA encod	350	18	54.2	1953	14	AEA56147	Aea56147 Streptoco
278	18	54.2	270	5	AAS76260	Aas76260 DNA encod	351	18	54.2	2000	8	ADA73448	Ada73448 Rice gene
279	18	54.2	270	5	AAS75573	Aas75573 DNA encod	352	18	54.2	2000	11	ACL35310	Acl35310 Rice stre
280	18	54.2	320	5	ABV47825	Abv47825 Human pro	c 353	18	54.2	2121	11	ACL27531	Acl27531 Rice abio
281	18	54.2	376	11	ADT94503	Adt94503 Colon can	354	18	54.2	2145	6	ABZ32245	Abz32245 Candida a
282	18	54.2	376	11	ADK40985	Adk40985 Human cDN	355	18	54.2	2198	8	ABS57375	Abs57375 cDNA enco
283	18	54.2	394	12	ADQ36974	Adq36974 Cell prol	356	18	54.2	2239	4	AAH78799	Aah78799 Human G-p
284	18	54.2	394	12	ADQ15644	Adq15644 Rice stre	357	18	54.2	2259	5	AAS70320	Aas70320 DNA encod
285	18	54.2	411	6	ABV97567	Abv97567 Human pan	c 358	18	54.2	2371	2	AAT99565	Aat99565 Human ago
286	18	54.2	432	6	ABQ59094	Abq59094 Human col	359	18	54.2	2501	14	ADV97941	Adv97941 Human tum
287	18	54.2	446	14	ADT76065	Adt76065 Human col	360	18	54.2	2742	5	AAH46102	Aah46102 Human CLC
288	18	54.2	453	9	ACH14280	Ach14280 Human adu	361	18	54.2	2742	9	ADA37185	Ada37185 Human CLC
289	18	54.2	460	14	ACL54478	Acl54478 Human col	362	18	54.2	2742	10	ADC51724	Adc51724 Human CLC
290	18	54.2	463	13	ADR60999	Adr60999 Cotton CD	363	18	54.2	2742	11	ADZ26326	Adz26326 Human rhl
291	18	54.2	481	11	ADT97273	Adt97273 Colon can	364	18	54.2	2742	12	ADI29549	Adi29549 Human CLC
292	18	54.2	481	11	ADX43755	Adx43755 Human cDN	365	18	54.2	2742	14	AEA35969	Aea35969 Novel tra
293	18	54.2	482	6	ABV97193	Abv97193 Human pan	366	18	54.2	2745	2	AAF81927	Aaf81927 Human ICA
294	18	54.2	482	11	ADT96218	Adt96218 Colon can	367	18	54.2	2763	4	AAF80429	Aaf80429 5-enolpyr
295	18	54.2	482	11	ADT94662	Adt94662 Colon can	368	18	54.2	2813	5	AEH28219	Aeh28219 Human CS1
296	18	54.2	482	11	ADT96722	Adt96722 Colon can	369	18	54.2	2825	15	AAH46124	Aah46124 Human CLC
297	18	54.2	482	11	ADT96379	Adt96379 Colon can	370	18	54.2	2854	4	AAF81787	Aaf81787 Human sec
298	18	54.2	482	11	ADX43204	Adx43204 Human cDN	371	18	54.2	2854	4	AAH34879	Aah34879 Human col
299	18	54.2	482	11	ADX42700	Adx42700 Human cDN	372	18	54.2	2867	4	AAH33285	Aah33285 Human col
300	18	54.2	482	11	ADX41144	Adx41144 Human cDN	373	18	54.2	2882	12	ADJ27235	Adj27235 Human TRI
301	18	54.2	482	11	ADX42861	Adx42861 Human cDN	374	18	54.2	2983	15	AEH28221	Aeh28221 Human CS1
302	18	54.2	483	11	ADT97275	Adt97275 Colon can	375	18	54.2	3007	6	AAI70801	Aai70801 Human cal
303	18	54.2	483	11	ADX43757	Adx43757 Human cDN	376	18	54.2	3007	9	ACH66411	Ach66411 Human cal
304	18	54.2	524	6	ABV89223	Abv89223 Human col	377	18	54.2	3007	14	ADW11215	Adw11215 Human cal
305	18	54.2	541	3	AAF08538	Aaf08538 Fumarium	378	18	54.2	3109	4	AAH35019	Aah35019 Human col
306	18	54.2	541	13	ADU52579	Adu52579 Fumarium	379	18	54.2	3111	2	AZ09840	Az09840 Human mem
307	18	54.2	541	14	ADZ90582	Adz90582 Fumarium	380	18	54.2	3111	9	ADA11073	Ada11073 Human cDN
308	18	54.2	675	13	ADQ60139	Adq60139 Human G p	381	18	54.2	3111	12	ADG47986	Adg47986 Human cal
309	18	54.2	876	12	ADQ58726	Adq58726 Human col	382	18	54.2	3186	8	ACC44827	Acc44827 Human G p
310	18	54.2	878	2	AAT45884	Aat45884 Human col	383	18	54.2	3186	15	AEB27527	Aeb27527 Human G p

384	18	54.2	3267	4	AAS26987	Aas26987 cDNA enco	C 457	17.8	53.6	196	6	ABS17468	AbS17468 Human gen
385	18	54.2	3267	10	ADB93165	AdB93165 Human cDN	C 458	17.8	53.6	282	6	ABL70829	AbL70829 Corn tass
386	18	54.2	3311	4	AAI29502	AaI29502 C902P det	C 459	17.8	53.6	294	10	ABX85996	AbX85996 Corn ear-
387	18	54.2	3311	8	ABZ33688	AbZ33688 Human col	C 460	17.8	53.6	297	8	ABZ22706	AbZ22706 Human ant
388	18	54.2	3311	10	AD63139	Ad63139 Human DNA	C 461	17.8	53.6	367	4	AAI14971	AaI14971 Probe #49
389	18	54.2	3311	10	AD63138	Ad63138 Human chl	C 462	17.8	53.6	367	4	ABA56707	AbA56707 Human foe
390	18	54.2	3311	10	AD63173	Ad63173 Human chl	C 463	17.8	53.6	367	4	AAI36319	AaI36319 Probe #50
391	18	54.2	3311	10	AD63174	Ad63174 Human DNA	C 464	17.8	53.6	367	4	ABA46158	AbA46158 Human bre
392	18	54.2	3311	10	AD62775	Ad62775 Human chl	C 465	17.8	53.6	367	4	ABA26319	AbA26319 Probe #47
393	18	54.2	3311	10	AD62774	Ad62774 Human chl	C 466	17.8	53.6	367	4	AAK30362	AaK30362 Human bon
394	18	54.2	3311	10	AD62739	Ad62739 Human chl	C 467	17.8	53.6	367	4	AAK04837	AaK04837 Human bra
395	18	54.2	3311	10	AD62740	Ad62740 Human DNA	C 468	17.8	53.6	367	4	ABS30003	AbS30003 Human liv
396	18	54.2	3311	11	ADP88234	AdP88234 Colon can	C 469	17.8	53.6	367	5	AAI04739	AaI04739 Probe #47
397	18	54.2	3311	11	ADP88270	AdP88270 Human dia	C 470	17.8	53.6	367	6	ABS04957	AbS04957 Human gen
398	18	54.2	3311	11	ADP88269	AdP88269 Colon can	C 471	17.8	53.6	375	9	ACH20498	AcH20498 Human adu
399	18	54.2	3311	11	ADP88235	AdP88235 Human dia	C 472	17.8	53.6	407	8	ABX48026	AbX48026 Bovine ES
400	18	54.2	3311	12	ADO78136	AdO78136 Human CLC	C 473	17.8	53.6	429	9	ACH46718	AcH46718 Human inf
401	18	54.2	3311	14	AEA04442	AeA04442 Human cDN	C 474	17.8	53.6	433	11	ADT96318	AdT96318 Colon can
402	18	54.2	3311	14	AEF53717	AeF53717 Human cal	C 475	17.8	53.6	433	11	ADX42800	AdX42800 Human cDN
403	18	54.2	3311	15	ABE48672	AbE48672 Human CLC	C 476	17.8	53.6	459	6	ABE93699	AbE93699 Arabidops
404	18	54.2	3311	8	ABZ09910	AbZ09910 Human 5'	C 477	17.8	53.6	513	4	AAH12665	AaH12665 Human cDN
405	18	54.2	4210	4	ABL07074	AbL07074 Drosophill	C 478	17.8	53.6	531	10	ADH13134	AdH13134 H influen
406	18	54.2	4569	6	ABA91304	AbA91304 Human col	C 479	17.8	53.6	531	10	ADH13136	AdH13136 H influen
407	18	54.2	7172	2	AAV52253	AaV52253 Streptoco	C 480	17.8	53.6	531	14	ABE10699	AbE10699 Haemophil
408	18	54.2	7411	10	ACA55370	AcA55370 Transform	C 481	17.8	53.6	531	14	ABE10697	AbE10697 Haemophil
409	18	54.2	10716	6	ABL33418	AbL33418 Human imm	C 482	17.8	53.6	633	11	ACL34102	AcL34102 Rice abio
410	18	54.2	17321	10	ADB81091	AdB81091 Human enz	C 483	17.8	53.6	645	12	ADI53010	AdI53010 Human chr
411	18	54.2	32892	14	AEA61109	AeA61109 Human CLC	C 484	17.8	53.6	651	14	ADW85541	AdW85541 MAP3K9 ma
412	18	54.2	36851	6	AAO28072	AaO28072 Human kin	C 485	17.8	53.6	688	4	AAH32451	AaH32451 Human olf
413	18	54.2	44990	9	ADO27111	AdO27111 Mouse Vda	C 486	17.8	53.6	717	8	ABT32194	AbT32194 Human neu
414	18	54.2	44990	10	ADB72449	AdB72449 Mouse Vda	C 487	17.8	53.6	728	4	AAH06438	AaH06438 Human cDN
415	18	54.2	44990	10	ADE95959	AdE95959 Mouse Vda	C 488	17.8	53.6	773	2	AAV34177	AaV34177 Human sec
416	18	54.2	57296	4	AAK78847	AaK78847 Human imm	C 489	17.8	53.6	773	8	ACD08048	AcD08048 cDNA enco
417	18	54.2	57296	4	AAK78170	AaK78170 Human imm	C 490	17.8	53.6	773	14	ADZ12234	AdZ12234 Human sec
418	18	54.2	57296	4	AAK79364	AaK79364 Human imm	C 491	17.8	53.6	911	13	ADX59374	AdX59374 Plant ful
419	18	54.2	57296	4	AAK86799	AaK86799 Human imm	C 492	17.8	53.6	958	13	ADX59371	AdX59371 Plant ful
420	18	54.2	61103	11	ACN43886	AcN43886 Human gen	C 493	17.8	53.6	959	13	ADX59365	AdX59365 Plant ful
421	18	54.2	77932	11	ADL27149	AdL27149 Mouse gen	C 494	17.8	53.6	1040	11	ACN86579	AcN86579 Breast ca
422	18	54.2	78028	9	ADA03071	AdA03071 Mouse mCG	C 495	17.8	53.6	1137	13	ADX30414	AdX30414 Plant ful
423	18	54.2	78028	9	ADA66355	AdA66355 Mouse mCG	C 496	17.8	53.6	1166	6	ABX15078	AbX15078 cDNA enco
424	18	54.2	78028	10	ADB72809	AdB72809 Mouse mCG	C 497	17.8	53.6	1182	4	AAU05641	AaU05641 Human rep
425	18	54.2	80276	14	ADZ12738	AdZ12738 Murine ca	C 498	17.8	53.6	1183	4	AAU05642	AaU05642 Human rep
426	18	54.2	98472	6	ABS54549	AbS54549 Human mul	C 499	17.8	53.6	1253	2	AAV34241	AaV34241 Human sec
427	18	54.2	98472	8	ACF62745	AcF62745 Cancer ba	C 500	17.8	53.6	1253	8	ACD08112	AcD08112 cDNA enco
428	18	54.2	98472	8	ADB20860	AdB20860 MRP1 base	C 501	17.8	53.6	1253	14	ADZ12298	AdZ12298 Human sec
429	18	54.2	98472	10	ADB87949	AdB87949 Human UGT	C 502	17.8	53.6	1352	4	ABU15899	AbU15899 Drosophill
430	18	54.2	98472	10	ADB96932	AdB96932 Human MDR	C 503	17.8	53.6	1428	3	ACG78073	AaC78073 Human can
431	18	54.2	98472	10	ADB92123	AdB92123 Human MDR	C 504	17.8	53.6	1460	4	AAZ23115	AaZ23115 DNA enco
432	18	54.2	110000	10	ABS56454_14	AbS56454_14 Continuation (15 o	C 505	17.8	53.6	1460	4	AAZ24928	AaZ24928 Human bre
433	18	54.2	110000	14	AEA61163_1	AcF62749 Cancer ba	C 506	17.8	53.6	1562	3	AAU77963	AaU77963 Human can
434	18	54.2	128993	8	ACF62749	AdB20868 MRP1 base	C 507	17.8	53.6	1584	9	ADB06347	AdB06347 Alloiococ
435	18	54.2	128993	8	ADB20868	AdB7957 Human UGT	C 508	17.8	53.6	1584	9	ADB06351	AdB06351 Alloiococ
436	18	54.2	128993	10	ADB87957	AdB96940 Human MDR	C 509	17.8	53.6	1584	9	ADB06345	AdB06345 Alloiococ
437	18	54.2	128993	10	ADB96940	AdB96940 Human MDR	C 510	17.8	53.6	1584	9	ADB06349	AdB06349 Alloiococ
438	18	54.2	128993	10	ADB92131	AdB92131 Human MDR	C 511	17.8	53.6	1615	13	ADX34982	AdX34982 Plant ful
439	18	54.2	137560	12	ADP74816	AdP74816 Parapoxvi	C 512	17.8	53.6	1724	14	ADZ64417	AdZ64417 Human can
440	18	54.2	147419	6	ABK83574	AbK83574 Human cDN	C 513	17.8	53.6	1776	14	ADV43710	AdV43710 Human psy
441	18	54.2	153752	12	ADO97531	AdO97531 Human gen	C 514	17.8	53.6	1861	15	AEF02361	AeF02361 Cotton cd
442	18	54.2	160921	11	ACN44962	AcN44962 Human gen	C 515	17.8	53.6	1984	4	ABU28309	AbU28309 Drosophill
443	18	54.2	178870	10	ADU13312	AdU13312 Osteoarth	C 516	17.8	53.6	1995	6	ABN95778	AbN95778 Gene #227
444	18	54.2	192992	13	ADB32866	AdB32866 Mouse can	C 517	17.8	53.6	2000	11	ACL38065	AcL38065 Rice stre
445	18	54.2	195102	13	ADV70461	AdV70461 Chicken g	C 518	17.8	53.6	2115	8	ACA29833	AcA29833 Prokaryot
446	18	54.2	275790	14	ADZ00150	AdZ00150 Salmonell	C 519	17.8	53.6	2247	13	ADR07715	AdR07715 Full leng
447	17.8	53.6	60	6	ABN47337	AbN47337 Human spl	C 520	17.8	53.6	2318	4	AAK53300	AaK53300 Human pol
448	17.8	53.6	196	4	AAI24159	AaI24159 Probe #14	C 521	17.8	53.6	2337	12	ABE26879	AbE26879 Pinus rad
449	17.8	53.6	196	4	ABA69281	AbA69281 Human foe	C 522	17.8	53.6	2493	14	ADO62064	AdO62064 Transcrip
450	17.8	53.6	196	4	AAI49448	AaI49448 Probe #18	C 523	17.8	53.6	2523	10	ADA533709	AdA533709 Human cod
451	17.8	53.6	196	4	ABA51270	AbA51270 Human bre	C 524	17.8	53.6	2623	4	AAK52316	AaK52316 Human pol
452	17.8	53.6	196	4	ABA36204	AbA36204 Probe #14	C 525	17.8	53.6	2638	4	ABL25138	AbL25138 Drosophill
453	17.8	53.6	196	4	AAK43379	AaK43379 Human bon	C 526	17.8	53.6	2641	8	ABX34705	AbX34705 Human mdd
454	17.8	53.6	196	4	AAK17568	AaK17568 Human bra	C 527	17.8	53.6	2644	10	ADC32066	AdC32066 Human nov
455	17.8	53.6	196	4	ABS43002	AbS43002 Human liv	C 528	17.8	53.6	2648	10	ADB62195	AdB62195 Human cDN
456	17.8	53.6	196	5	AAI09724	AaI09724 Probe #97	C 529	17.8	53.6	2676	11	ACL28094	AcL28094 Rice abio

530	17.8	53.6	2851	14	ADZ70724	Chromosom	Adz70724	603	17.8	53.6	160820	8	ABQ76673	Androgen
531	17.8	53.6	3057	15	Aee86935	Human che	Aee86935	604	17.8	53.6	196063	13	ABD33462	Human can
532	17.8	53.6	3122	12	ADQ63658	Novel hum	Adq63658	c 605	17.8	53.6	201766	10	ADL13771	Osteoarth
533	17.8	53.6	3144	4	ABL21333	Drosophil	AbL21333	c 606	17.8	53.6	212231	11	ACN44598	Human gen
534	17.8	53.6	3234	8	ABZ76275	Notch-int	Abz76275	c 607	17.8	53.6	272022	12	ADQ97126	Human can
535	17.8	53.6	3240	4	AAH14453	Human cDN	Aah14453	c 608	17.8	53.6	304905	11	ADP75180	Human End
536	17.8	53.6	3422	4	AAH16210	Human cDN	Aah16210	c 609	17.8	53.6	349980	13	ADT05648	Haemophil
c 537	17.8	53.6	3724	4	ABL15898	Drosophil	AbL15898	610	17.6	53.0	40	12	ADI28527	Human GPC
538	17.8	53.6	3935	4	ABL06648	Drosophil	AbL06648	c 611	17.6	53.0	51	4	AAA89199	DNA encod
539	17.8	53.6	4083	5	ABV29639	Human pro	Abv29639	c 612	17.6	53.0	56	12	ADJ66760	Human CA1
540	17.8	53.6	4083	5	ABV23765	Human pro	Abv23765	613	17.6	53.0	186	10	ACA55771	Mouse sig
c 541	17.8	53.6	4094	14	ADZ64409	Human can	Adz64409	614	17.6	53.0	186	12	ADI55567	Human pol
542	17.8	53.6	4324	3	AAC60909	Human gen	Aac60909	c 615	17.6	53.0	260	10	ABX83902	Corn ear-
c 543	17.8	53.6	4960	4	ABL18720	Drosophil	AbL18720	c 616	17.6	53.0	273	10	ABX85253	Corn ear-
544	17.8	53.6	5279	3	RAC60913	Human squ	Rac60913	c 617	17.6	53.0	276	5	ABV07935	Human pro
545	17.8	53.6	5349	6	ABA02537	Immune re	AbA02537	c 618	17.6	53.0	296	10	ABX88439	Corn ear-
546	17.8	53.6	5412	6	ABA02536	Immune re	AbA02536	c 619	17.6	53.0	345	2	AAQ60107	Human bra
547	17.8	53.6	5641	4	ABL21332	Drosophil	AbL21332	c 620	17.6	53.0	369	5	AAQ65648	DNA encod
548	17.8	53.6	5757	8	ABX63349	Human cDN	Abx63349	621	17.6	53.0	369	5	ABV37858	Human pro
549	17.8	53.6	5833	6	ABA02535	Immune re	AbA02535	622	17.6	53.0	372	6	ABQ97774	Mouse ES
550	17.8	53.6	5965	6	ABV73788	Human pap	Abv73788	c 623	17.6	53.0	379	4	AAI83788	Human pol
551	17.8	53.6	5965	9	ACD26736	Papilloma	AcD26736	c 624	17.6	53.0	388	4	AAK57060	Human imm
552	17.8	53.6	6028	6	ABA02534	Immune re	AbA02534	625	17.6	53.0	393	3	AAA35154	Human ade
553	17.8	53.6	6028	6	ABV73789	Human pap	Abv73789	626	17.6	53.0	393	3	AAF21276	Human low
554	17.8	53.6	6028	9	ACD26737	Papilloma	AcD26737	627	17.6	53.0	393	10	ABZ96970	Human nuc
555	17.8	53.6	6028	10	ADF09636	Human sig	Adf09636	628	17.6	53.0	393	11	ABD20819	Human pul
556	17.8	53.6	6028	12	ADL49290	Human PRO	AdL49290	629	17.6	53.0	469	13	ACF85420	Human SIR
557	17.8	53.6	6028	12	ADQ19450	Human sof	Adq19450	630	17.6	53.0	527	13	ACN58144	Cotton gy
558	17.8	53.6	6028	13	ADP25355	PRO polyP	Adp25355	631	17.6	53.0	529	4	ABA60419	Human foe
559	17.8	53.6	6028	15	AEE88052	Human cDN	Aee88052	632	17.6	53.0	529	4	RAI140303	Probe #89
560	17.8	53.6	6099	12	ADQ23664	Human sof	Adq23664	633	17.6	53.0	529	4	AAK34584	Human bon
561	17.8	53.6	7599	13	ADR4331	Aspergill	Adr4331	634	17.6	53.0	529	4	AAK08697	Human bra
c 562	17.8	53.6	7760	4	ABL15894	Drosophil	AbL15894	635	17.6	53.0	529	4	ABS34357	Human liv
563	17.8	53.6	7819	10	ADF81550	Leukaemia	Adf81550	636	17.6	53.0	595	13	ACN59327	Cotton gy
c 564	17.8	53.6	7947	4	ABL28308	Drosophil	AbL28308	637	17.6	53.0	672	4	AAF89964	Partial n
565	17.8	53.6	8062	8	ABX34538	Human mdd	Abx34538	638	17.6	53.0	700	4	AAH92725	Human inf
566	17.8	53.6	8070	13	ACN40496	Tumour-as	Acn40496	639	17.6	53.0	700	4	AAH92726	Human inf
567	17.8	53.6	8091	2	AAV57001	Human Not	Aav57001	640	17.6	53.0	706	3	AAF13054	Aspergill
568	17.8	53.6	8091	6	ABZ34974	Human gen	Abz34974	641	17.6	53.0	706	13	ADU57095	Aspergill
569	17.8	53.6	8091	10	ADL13816	Osteoarth	AdL13816	642	17.6	53.0	706	14	ADZ95098	Aspergill
570	17.8	53.6	8091	12	ADH34620	Notch hom	Adh34620	643	17.6	53.0	710	12	ADJ66748	Muc 16 St
571	17.8	53.6	8091	12	ADJ75112	Marker ge	Adj75112	644	17.6	53.0	731	3	AAC52525	Arabidops
572	17.8	53.6	8091	12	ADNA48549	Human Not	Adn48549	c 645	17.6	53.0	732	12	ADJ42949	Plant cDN
573	17.8	53.6	8091	13	ADP23495	PRO polyP	Adp23495	c 646	17.6	53.0	787	13	ADR61608	Cotton CD
574	17.8	53.6	8091	13	ADR83501	Human Not	Adr83501	c 647	17.6	53.0	788	14	AEB66160	Rice geno
575	17.8	53.6	8091	14	ADR70404	Human Not	Adr70404	c 648	17.6	53.0	810	12	ADJ42948	Plant cDN
576	17.8	53.6	8257	6	ABS70408	Human bon	Abs70408	649	17.6	53.0	894	8	AAI94653	Human neu
c 577	17.8	53.6	13295	4	ABL09488	Drosophil	AbL09488	650	17.6	53.0	894	8	ACA42823	Prokaryot
578	17.8	53.6	13340	2	AAZ23522	O. longis	Aaz23522	c 651	17.6	53.0	1346	2	AAZ97946	Human sec
c 579	17.8	53.6	15743	6	ABK28395	DNA trans	Abk28395	c 652	17.6	53.0	1346	9	ADA11561	Human cDN
580	17.8	53.6	17897	10	ADB54177	Pretrate	Adb54177	c 653	17.6	53.0	1346	13	ADX15317	Plant ful
c 581	17.8	53.6	17897	13	ADS89327	Oligonuc1	Ads89327	654	17.6	53.0	1620	11	ACL27597	Rice adio
c 582	17.8	53.6	17897	15	AEP08746	Human NDR	Aep08746	655	17.6	53.0	1620	12	ADJ39701	Plant cDN
c 583	17.8	53.6	18753	4	ABL20682	Drosophil	AbL20682	656	17.6	53.0	1731	10	ACF71816	Photorthab
c 584	17.8	53.6	21354	3	AAAS46815	Tumour su	Aas46815	c 657	17.6	53.0	1760	13	ADX49274	Plant ful
585	17.8	53.6	27425	3	RAAS53978	Dmp53 tum	Raa53978	658	17.6	53.0	1958	6	AAD28059	Human ade
c 586	17.8	53.6	32167	5	ABA20621	Human ner	AbA20621	c 589	17.6	53.0	2223	10	ABZ40838	N. gonorr
c 587	17.8	53.6	40568	11	ACN44896	Mouse gen	Acn44896	c 660	17.6	53.0	2258	5	AAZ92412	DNA encod
c 588	17.8	53.6	41150	10	ADL13819	Osteoarth	AdL13819	661	17.6	53.0	2266	4	AAK71354	Human imm
c 589	17.8	53.6	41150	14	AED18343	Fibrotic	Aed18343	c 662	17.6	53.0	2304	13	ADX60952	Plant ful
590	17.8	53.6	44348	12	ADNA48556	Human Not	Adn48556	c 663	17.6	53.0	2389	6	ABQ70840	Listeria
591	17.8	53.6	48727	4	AAK67375	Human imm	Aak67375	664	17.6	53.0	2403	2	AAZ33603	Human bre
c 592	17.8	53.6	52464	14	ADZ12858	Murine ca	Adz12858	c 665	17.6	53.0	2823	15	AEP08723	Human gen
c 593	17.8	53.6	52764	11	ACN43932	Mouse gen	Acn43932	c 666	17.6	53.0	2823	15	AEP55067	Human gen
c 594	17.8	53.6	70123	11	ACN44254	Human gen	Acn44254	667	17.6	53.0	3108	11	ACN44331	Human mRN
595	17.8	53.6	96594	9	ADA042756	Human MOR	Ada042756	668	17.6	53.0	3126	8	ACA26007	Prokaryot
596	17.8	53.6	96594	10	ADB72494	Human MOR	Adb72494	669	17.6	53.0	3169	3	AAA35156	Human ade
597	17.8	53.6	96594	10	ADC85236	Human MOR	Adc85236	670	17.6	53.0	3169	3	AAF21278	Human low
598	17.8	53.6	96594	12	ADM74351	Human cor	Adm74351	671	17.6	53.0	3169	10	ABZ96972	Human nuc
c 599	17.8	53.6	110000	2	AAT42063_01	Continuation (2 of	Aat42063_01	672	17.6	53.0	3169	11	ABD20821	Human pul
c 600	17.8	53.6	110000	9	ADB12064_00	Allioococ	Adb12064_00	673	17.6	53.0	3446	12	ADO60053	CRH signa
c 601	17.8	53.6	110000	14	AEA61160_1	Continuation (2 of	Aea61160_1	674	17.6	53.0	3561	2	AAQ55970	PTase PT
602	17.8	53.6	110079	13	ADV35020	Murine cD	Adv35020	675	17.6	53.0	3566	6	AB199237	Mouse isc

676	17.6	53.0	3748	6	ABK34630	Abk34630 Human cDN	749	17.4	52.4	364	4	AA118044	Aa118044 Human bre
677	17.6	53.0	3804	4	AAH16044	Aah16044 Human cDN	750	17.4	52.4	415	11	ACN92660	Acn92660 Breast ca
678	17.6	53.0	4020	6	ABQ54212	Abq54212 Human ova	751	17.4	52.4	420	6	ABL78873	Ab178873 Human ova
679	17.6	53.0	4047	13	ADP24470	Adp24470 PRO polyP	752	17.4	52.4	457	13	ACF84030	Acf84030 Human SIR
680	17.6	53.0	4047	15	AE883887	Aee83887 Human cDN	753	17.4	52.4	489	5	AAH81814	Aah81814 Human dif
681	17.6	53.0	4091	4	AA158457	Aai58457 Human pol	754	17.4	52.4	489	5	AAH81638	Aah81638 Human dif
682	17.6	53.0	4091	5	ADQ98669	Adq98669 DNA encod	755	17.4	52.4	511	3	AAA43724	Aaa43724 Mouse sec
683	17.6	53.0	4091	9	ADB48429	Adb48429 Novel hum	756	17.4	52.4	516	15	AEF78593	Aef78593 Mouse gen
684	17.6	53.0	4093	4	AAI60243	Aai60243 Human pol	757	17.4	52.4	518	9	ACH28318	Ach28318 Human adu
685	17.6	53.0	4115	4	AAO20404	Aao20404 Human sec	758	17.4	52.4	526	12	ADN13627	Adn13627 Human pro
686	17.6	53.0	4115	8	ABZ73366	Abz73366 Secreted	759	17.4	52.4	533	6	ABQ36864	Abq36864 Oligonucl
687	17.6	53.0	4115	10	ADC20115	Adc20115 Human sec	760	17.4	52.4	532	6	ABQ36865	Abq36865 Oligonucl
688	17.6	53.0	4115	10	ABT16788	Abt16788 Human sec	761	17.4	52.4	546	11	ACN86819	Acn86819 Breast ca
689	17.6	53.0	4115	10	ABQ66974	Abq66974 Human sec	762	17.4	52.4	555	5	ABV53410	Abv53410 Human pro
690	17.6	53.0	4128	10	ADF81727	Adf81727 Leukaemia	763	17.4	52.4	586	4	AA107903	Aal07903 Human bre
691	17.6	53.0	4290	5	ABL52917	Ab152917 2-keto-D-	764	17.4	52.4	618	6	ABK33035	Abk33035 DNA encod
692	17.6	53.0	4514	10	ADB56172	Adb56172 Toxicity-	765	17.4	52.4	618	12	ADQ92260	Adq92260 Human aut
693	17.6	53.0	4514	10	ADB53674	Adb53674 Primary r	766	17.4	52.4	621	3	AAC42900	Aac42900 Arabidops
694	17.6	53.0	4514	14	AEC11769	Aeci1769 Rat surro	767	17.4	52.4	652	5	AAS05408	Aas05408 Mammalian
695	17.6	53.0	4668	3	AACT77483	Aac77483 Human ORF	768	17.4	52.4	828	6	ABK74586	Abk74586 Bacillus
696	17.6	53.0	4675	14	ADY18887	Ady18887 DNA encod	769	17.4	52.4	845	5	AAS87938	Aas87938 DNA encod
697	17.6	53.0	5142	10	ADF90720	Adf90720 Human hep	770	17.4	52.4	927	11	ACH98632	Ach98632 Klebsiell
698	17.6	53.0	5142	12	ADN05411	Adn05411 Antipsori-	771	17.4	52.4	955	6	ABN98556	Abn98556 Arabidops
699	17.6	53.0	5142	13	ADP55825	Adp55825 Human PRO	772	17.4	52.4	955	10	ADD35246	Add35246 Mouse mit
700	17.6	53.0	7095	14	ADY58452	Ady58452 Leishmani	773	17.4	52.4	1008	8	ACA30213	Aca30213 Prokaryot
701	17.6	53.0	7565	12	ADL70703	Adl70703 Murine CR	774	17.4	52.4	1143	3	AAA59157	Aaa59157 cDNA enco
702	17.6	53.0	7657	4	AAAS45478	Aas45478 Chemically	775	17.4	52.4	1143	3	AAA58985	Aaa58985 cDNA enco
703	17.6	53.0	7657	6	ABL34023	Ab134023 Human imm	776	17.4	52.4	1143	10	ADB31999	Adb31999 Mouse cDN
704	17.6	53.0	14328	4	AAK71326	Aak71326 Human imm	777	17.4	52.4	1143	12	ADR13801	Adr13801 Murine in
705	17.6	53.0	14328	4	AAK71325	Aak71325 Human imm	778	17.4	52.4	1215	13	ADS46182	Ads46182 Bacterial
706	17.6	53.0	14328	4	AAK71335	Aak71335 Human imm	779	17.4	52.4	1221	3	AZ521198	Aaz521198 Murine tr
707	17.6	53.0	14328	4	AAK71356	Aak71356 Human imm	780	17.4	52.4	1221	10	AAD54508	Aad54508 Mouse int
708	17.6	53.0	15000	3	ABN97977	Abn97977 Human ret	781	17.4	52.4	1221	12	ADQ92005	Adq92005 Mouse tra
709	17.6	53.0	16878	3	AAA81515	Aaa81515 N. mening	782	17.4	52.4	1228	13	ADT17280	Adt17280 Plant cDN
710	17.6	53.0	21234	6	AAD32039	Aad32039 Human kin	783	17.4	52.4	1258	4	AAD10108	Aad10108 Mouse zCY
711	17.6	53.0	23615	14	AEC09537	Aeco9537 Human Bli	784	17.4	52.4	1269	8	ACF75039	Acf75039 Staphyloc
712	17.6	53.0	25837	4	AAK59592	Aak59592 Human imm	785	17.4	52.4	1284	6	ABN90749	Abn90749 Staphyloc
713	17.6	53.0	28983	14	AE9B96521	Aeb96521 Human ADA	786	17.4	52.4	1284	13	ADS02816	Ads02816 Staphyloc
714	17.6	53.0	36741	4	AAAS03388	Aas03388 Human ade	787	17.4	52.4	1318	3	AAC48785	Aac48785 Arabidops
715	17.6	53.0	36741	6	ABQ88153	Abq88153 Human ost	788	17.4	52.4	1319	3	AAC38564	Aac38564 Arabidops
716	17.6	53.0	36741	6	ABS69890	Abs69890 Human ade	789	17.4	52.4	1356	10	ADB69774	Adb69774 C. neofor
717	17.6	53.0	36741	6	ABS52806	Abs52806 DNA encod	790	17.4	52.4	1425	3	AAC38512	Aac38512 Arabidops
718	17.6	53.0	36741	15	AEF10051	Aef10051 Human ade	791	17.4	52.4	1425	3	AAC48792	Aac48792 Arabidops
719	17.6	53.0	38920	12	ADQ97252	Adq97252 Mouse can	792	17.4	52.4	1469	13	ADT14803	Adt14803 Plant cDN
720	17.6	53.0	39325	4	AAK81660	Aak81660 Human imm	793	17.4	52.4	1505	10	ADB69413	Adb69413 C. neofor
721	17.6	53.0	40046	6	ABL68361	Ab168361 Kidney ca	794	17.4	52.4	1545	2	AAT90510	Aat90510 DNA encod
722	17.6	53.0	40046	6	ABL68822	Ab168822 Kidney ca	795	17.4	52.4	1559	3	AAC48790	Aac48790 Arabidops
723	17.6	53.0	40046	6	ABL68140	Ab168140 Kidney ca	796	17.4	52.4	1562	3	AAC38506	Aac38506 Arabidops
724	17.6	53.0	41685	11	ACN44330	Acn44330 Human gen	797	17.4	52.4	1562	3	ABL61903	Ab161903 Colon ade
725	17.6	53.0	42123	11	ACN44306	Acn44306 Human gen	798	17.4	52.4	1592	6	ABL64093	Ab164093 Breast ca
726	17.6	53.0	53131	14	AEA02956	Aea02956 Human pai	799	17.4	52.4	1592	6	ABT11013	Abt11013 Human bre
727	17.6	53.0	54550	9	ACF05021	Acf05021 Human cat	800	17.4	52.4	1659	13	ADT7784	Adt7784 Plant cDN
728	17.6	53.0	58922	13	ABD33407	Abd33407 Human can	801	17.4	52.4	1674	10	ACC61917	Acc61917 Gene sequ
729	17.6	53.0	62154	12	ADO97611	Ado97611 Human can	802	17.4	52.4	1674	10	ADK64875	Adk64875 Disease t
730	17.6	53.0	69622	13	ABD333115	Abd333115 Human can	803	17.4	52.4	1674	12	ADJ64177	Adj64177 S. cerevi
731	17.6	53.0	79860	11	ACN44122	Acn44122 Human gen	804	17.4	52.4	1692	5	AAS80345	Aas80345 DNA encod
732	17.6	53.0	105184	6	ABK24122	Abk24122 Bacterial	805	17.4	52.4	1692	5	AAS69033	Aas69033 DNA encod
733	17.6	53.0	110000	3	AAA81490	Aaa81490 Continuation (10 o	806	17.4	52.4	1710	8	ADA53951	Ada53951 Human cod
734	17.6	53.0	110000	10	ACF67367_50	Acf67367_50 Continuation (51 o	807	17.4	52.4	1728	10	ACA34394	ACA34394 Prokaryot
735	17.6	53.0	110000	10	ACF65387_1	Acf65387_1 Continuation (2 of	808	17.4	52.4	1897	11	ADN95347	Adn95347 Human BEC
736	17.6	53.0	110000	11	ADN27081_01	Adn27081_01 Continuation (2 of	809	17.4	52.4	1917	13	ADS78310	Ads78310 Aminotran
737	17.6	53.0	113575	11	ACN44348	Acn44348 Mouse gen	810	17.4	52.4	1921	13	AA126706	Aal26706 Human bre
738	17.6	53.0	147708	6	ABQ88154	Abq88154 Human ost	811	17.4	52.4	1928	11	ACN88684	Acn88684 Breast ca
739	17.6	53.0	170001	14	AED18493	Aed18493 Fibrotic	812	17.4	52.4	1981	5	AAS83523	Aas83523 DNA encod
740	17.6	53.0	176771	12	ADQ97167	Adq97167 Human can	813	17.4	52.4	2000	11	ACL35768	Acl35768 Rice stre
741	17.6	53.0	193672	10	ADL13570	Adl13570 Osteoarth	814	17.4	52.4	2079	2	AAT90513	Aat90513 DNA encod
742	17.6	53.0	209484	11	ACN44126	Acn44126 Human gen	815	17.4	52.4	2143	5	AAS84876	Aas84876 DNA encod
743	17.6	53.0	349980	3	AAF21609	Aaf21609 Neisseria	816	17.4	52.4	2212	8	AAS53176	Aas53176 Human kin
744	17.6	53.0	349980	3	AAF21608	Aaf21608 Neisseria	817	17.4	52.4	2427	5	AAS69327	Aas69327 DNA encod
745	17.4	52.4	295	3	AAA87092	Aaa87092 Rat hepat	818	17.4	52.4	2569	10	ADD44980	Add44980 Rat gene
746	17.4	52.4	315	10	ACD95013	Acd95013 Human col	819	17.4	52.4	2569	10	ADD44984	Add44984 Rat gene
747	17.4	52.4	350	4	AA125649	Aal25649 Human bre	820	17.4	52.4	2697	4	ABL07429	Ab107429 Drosophil
748	17.4	52.4	350	6	ABV96297	Abv96297 Human pan	821	17.4	52.4	2876	11	ADM02841	Adm02841 Human cDN

C 822	17.4	52.4	2876	14	AEC85771	Aec85771 Human cDN	C 895	17.4	52.4	165799	14	ADZ13004	Adz13004 Murine ca
C 823	17.4	52.4	2889	9	ABL16704	Ab116704 Drosophil	C 896	17.4	52.4	171162	14	AED89398	Aed89398 Human bre
C 824	17.4	52.4	2932	9	ADN97572	Adn97572 Rat presy	C 897	17.4	52.4	194266	14	AED76144	Aed76144 Mouse CA
825	17.4	52.4	3112	10	ADB58545	Adb58545 Toxicity-	C 898	17.4	52.4	235033	2	AAV57926	Aav57926 Hereditar
826	17.4	52.4	3305	10	ADD29569	Add29569 Human tum	C 899	17.4	52.4	236303	4	AAV51614	Aav51614 Human gen
827	17.4	52.4	3305	10	ADD29624	Add29624 Human tum	C 900	17.4	52.4	237326	2	AAV57903	Aav57903 Hereditar
C 828	17.4	52.4	3306	3	AAA877701	Aaa877701 Human sec	C 901	17.4	52.4	246386	14	AEE05008	Aee05008 Cancer-as
C 829	17.4	52.4	3306	8	ADA40549	Ada40549 Human sec	C 902	17.4	52.4	289190	13	ABD33143	Abd33143 Murine ca
C 830	17.4	52.4	3306	8	ACC50852	Acc50852 Human sec	C 903	17.4	52.4	349980	13	ADT05649	Adt05649 Haemophil
C 831	17.4	52.4	3306	8	ABZ71477	Abz71477 Secreted	C 904	17.2	51.8	49	1	AAH30053	Aah30053 Sequence
C 832	17.4	52.4	3306	9	ADB91444	Adb91444 Human sec	C 905	17.2	51.8	213	5	ABA14740	Abal4740 Human ner
C 833	17.4	52.4	3306	10	ADC73984	Adc73984 Human sec	C 906	17.2	51.8	275	12	ADP91609	Adp91609 Cotton ex
C 834	17.4	52.4	3306	10	ADA56695	Ada56695 Gene enco	C 907	17.2	51.8	282	10	ABX82669	Abx82669 Corn ear-
C 835	17.4	52.4	3308	3	AAA87681	Aaa87681 Human sec	C 908	17.2	51.8	305	12	ADP94439	Adp94439 Cotton ex
C 836	17.4	52.4	3308	8	ADA40228	Ada40228 Human sec	C 909	17.2	51.8	331	13	ACN55724	Acn55724 Cotton an
C 837	17.4	52.4	3308	8	ACC50671	Acc50671 Human sec	C 910	17.2	51.8	342	12	ADP91778	Adp91778 Cotton ex
C 838	17.4	52.4	3308	8	ABZ71365	Abz71365 Secreted	C 911	17.2	51.8	353	5	ABV00246	Abv00246 Human pro
C 839	17.4	52.4	3308	9	ADB91306	Adb91306 Human sec	C 912	17.2	51.8	362	5	ABV09415	Abv09415 Human pro
C 840	17.4	52.4	3308	10	ADC73754	Adc73754 Human sec	C 913	17.2	51.8	363	12	ADP94898	Adp94898 Cotton ex
C 841	17.4	52.4	3308	10	ADA56388	Ada56388 Gene enco	C 914	17.2	51.8	364	12	ADP95566	Adp95566 Cotton ex
C 842	17.4	52.4	3447	13	ADX33598	Adx33598 Plant ful	C 915	17.2	51.8	376	12	ADP91503	Adp91503 Cotton ex
C 843	17.4	52.4	3505	10	ADB69052	Adb69052 C. neofor	C 916	17.2	51.8	393	12	ADP94805	Adp94805 Cotton ex
844	17.4	52.4	3516	2	AAV74546	Aav74546 Staphyloc	C 917	17.2	51.8	396	4	AAF94885	Aaf94885 Human ova
845	17.4	52.4	3524	4	ABL07298	Ab107298 Drosophil	C 918	17.2	51.8	396	6	ABL48835	Ab148835 Ovarian c
846	17.4	52.4	3899	6	ABI99205	Abi99205 Mouse isc	C 919	17.2	51.8	396	6	ABT03152	Abt03152 Human ova
847	17.4	52.4	5048	5	AA871638	Aas71638 DNA enco	C 920	17.2	51.8	396	11	ADM10745	Adm10745 Human ova
C 848	17.4	52.4	5174	13	ADT05447	Adt05447 Haemophil	C 921	17.2	51.8	396	12	ADJ11075	Adj11075 Represent
849	17.4	52.4	5443	14	ADM94163	Adw94163 Staphyloc	C 922	17.2	51.8	396	12	ADM43336	Adm43336 Human ova
850	17.4	52.4	5565	14	ADM44323	Adw14323 Human myo	C 923	17.2	51.8	405	4	AA838582	Aas38582 Novel hum
851	17.4	52.4	5885	15	AEF18698	Aef18698 Human KIA	C 924	17.2	51.8	405	5	AA892050	Aas92050 DNA enco
C 852	17.4	52.4	6000	13	ADY81055	Ady81055 Rice anti	C 925	17.2	51.8	406	5	ABV39560	Abv39560 Human pro
853	17.4	52.4	6126	10	ADC13546	Adc13546 Human NOV	C 926	17.2	51.8	406	5	ABV30591	Abv30591 Human pro
854	17.4	52.4	6151	6	ABZ11646	Abz11646 Human pol	C 927	17.2	51.8	409	10	ADF80394	Adf80394 Leukaemia
855	17.4	52.4	6151	12	ADM44164	Adm44164 Novel hum	C 928	17.2	51.8	418	6	ABL63023	Ab163023 Breast ca
856	17.4	52.4	6693	4	ABL07428	Ab107428 Drosophil	C 929	17.2	51.8	418	6	ABL63244	Ab163244 Breast ca
C 857	17.4	52.4	6881	6	ABL33381	Ab133381 Human imm	C 930	17.2	51.8	418	6	ABL62618	Ab162618 Colon ade
C 858	17.4	52.4	6881	6	ABK28250	Abk28250 DNA trans	C 931	17.2	51.8	418	6	ABN94919	Abn94919 Gene #141
C 859	17.4	52.4	8280	3	AAA95949	Aaa95949 Human KIK	C 932	17.2	51.8	443	12	ADP95793	Adp95793 Cotton ex
C 860	17.4	52.4	8280	12	ADN10931	Adn10931 Human kal	C 933	17.2	51.8	445	4	AAI80578	Aai80578 Human pol
C 861	17.4	52.4	10848	12	ADL81731	Adl81731 P. aerugi	C 934	17.2	51.8	447	3	AA869716	Aac69716 Human bre
862	17.4	52.4	12551	4	ABL9156	Ab19156 Drosophil	C 935	17.2	51.8	450	4	AA838533	Aas38533 Novel hum
863	17.4	52.4	14766	4	ABL05848	Ab105848 Drosophil	C 936	17.2	51.8	450	14	ADV75468	Adv75468 Human col
C 864	17.4	52.4	19338	14	AEA61189	Aea61189 Human DNA	C 937	17.2	51.8	514	2	AAV75344	Aav75344 Staphyloc
C 865	17.4	52.4	23377	11	ACN44808	Acn44808 Mouse gen	C 938	17.2	51.8	517	10	ADF80417	Adf80417 Leukaemia
C 866	17.4	52.4	25599	4	ABL16668	Ab116668 Drosophil	C 939	17.2	51.8	529	13	ACN50568	Acn50568 Cotton ma
C 867	17.4	52.4	35510	13	ADM39407	Adw39407 Herpes si	C 940	17.2	51.8	539	13	ADQ49450	Adq49450 Novel can
C 868	17.4	52.4	44400	4	AA812438	Aae12438 DNA enco	C 941	17.2	51.8	539	13	ACN48426	Acn48426 Cotton pr
C 869	17.4	52.4	46275	6	ABT10145	Abt10145 Human bre	C 942	17.2	51.8	542	8	ABZ17596	Abz17596 S2 subtra
C 870	17.4	52.4	46275	10	ADL13621	Adl13621 Osteoarth	C 943	17.2	51.8	549	12	ACH75752	Ach75752 Human gen
C 871	17.4	52.4	49826	11	ACN44340	Acn44340 Mouse gen	C 944	17.2	51.8	553	12	ADP93149	Adp93149 Cotton ex
C 872	17.4	52.4	60327	9	ADA02681	Ada02681 Mouse Dus	C 945	17.2	51.8	562	12	ACH71075	Ach71075 Human gen
C 873	17.4	52.4	60327	10	ADB72419	Adb72419 Mouse Dus	C 946	17.2	51.8	563	10	ADK59431	Adk59431 Plant DNA
C 874	17.4	52.4	60327	10	ADE95929	Ade95929 Mouse Dus	C 947	17.2	51.8	564	13	ADQ50608	Adq50608 Novel can
C 875	17.4	52.4	81679	12	ADQ59189	Adq59189 MSI-H car	C 948	17.2	51.8	569	5	ABV53362	Abv53362 Human pro
C 876	17.4	52.4	83698	6	ABN85767	Abn85767 Arabidops	C 949	17.2	51.8	600	6	ABQ32874	Abq32874 Oligonuc1
C 877	17.4	52.4	89873	13	ABD32846	Abd32846 Mouse can	C 950	17.2	51.8	600	6	ABQ32875	Abq32875 Oligonuc1
C 878	17.4	52.4	91080	6	ABX08336_16	Continuation (17 o	C 951	17.2	51.8	603	13	ACN57282	Acn57282 Cotton gy
C 879	17.4	52.4	91134	12	ADO50281_16	Continuation (17 o	C 952	17.2	51.8	618	3	AA833701	Aac33701 Arabidops
C 880	17.4	52.4	91138	12	ADN97989_16	Continuation (17 o	C 953	17.2	51.8	649	6	ABT09926	Abt09926 Human bre
C 881	17.4	52.4	91139	12	ADJ25985_16	Continuation (17 o	C 954	17.2	51.8	653	10	ADK76605	Adk76605 DNA homol
C 882	17.4	52.4	91140	14	ABE85185_16	Continuation (17 o	C 955	17.2	51.8	653	10	ADK59249	Adk59249 Plant DNA
C 883	17.4	52.4	110000	2	AAT42063_12	Continuation (13 o	C 956	17.2	51.8	660	4	AA846942	Aas46942 Human G p
884	17.4	52.4	110000	6	ABQ69245_05	Continuation (6 of	C 957	17.2	51.8	660	6	ABK81720	Abk81720 CDNA enco
885	17.4	52.4	110000	6	ABQ69245_06	Continuation (7 of	C 958	17.2	51.8	727	4	AAI96953	Aai96953 Human neu
C 886	17.4	52.4	110000	6	ABQ67197_05	Continuation (6 of	C 959	17.2	51.8	751	6	AAD27742	Aad27742 Human ful
C 887	17.4	52.4	127369	11	ACN44006	Acn44006 Human gen	C 960	17.2	51.8	751	10	ACF65560	Acf65560 Photorhab
C 888	17.4	52.4	127943	12	ADQ97651	Adq97651 Human can	C 961	17.2	51.8	753	5	AA879897	Aas79897 DNA enco
C 889	17.4	52.4	135928	15	AEF75216	Aef75216 Human can	C 962	17.2	51.8	761	4	AAI95235	Aai95235 Human neu
C 890	17.4	52.4	150481	14	AED89397	Aed89397 Human bre	C 963	17.2	51.8	761	4	AAI98132	Aai98132 Human neu
C 891	17.4	52.4	152261	14	AED07415	Aed07415 DNA of HS	C 964	17.2	51.8	794	4	AAH08612	Aah08612 Human cDN
C 892	17.4	52.4	152261	14	AED00550	Aed00550 Herpes so	C 965	17.2	51.8	806	3	AA833981	Aac33981 Arabidops
C 893	17.4	52.4	152261	14	AED00550	Aed00550 Herpes so	C 966	17.2	51.8	852	4	AAH07577	Aah07577 Human cDN
C 894	17.4	52.4	165221	11	ACN44524	Acn44524 Mouse gen	C 967	17.2	51.8	861	4	AAH04431	Aah04431 Human cDN

968 17.2 51.8 883 10 ADC86650 Human GPC  
c 969 17.2 51.8 903 5 AAS81735  
Ado28921 Mouse nov  
970 17.2 51.8 918 12 AD028921  
Adt15659 Plant cdn  
971 17.2 51.8 947 13 ADT15659  
Acf72582 Staphyloc  
972 17.2 51.8 972 8 ACF72582  
Adx50470 Plant ful  
973 17.2 51.8 1020 13 ADX50470  
c 974 17.2 51.8 1057 14 AEB66373  
Aeb66373 Rice geno  
c 975 17.2 51.8 1072 11 ACN81830  
Acn81830 Breast ca  
c 976 17.2 51.8 1182 10 ADC71369  
Adc71369 Human col  
c 977 17.2 51.8 1282 10 ADE79051  
Ade79051 Human pro  
c 978 17.2 51.8 1296 10 ADE79038  
Ade79038 Human pro  
c 979 17.2 51.8 1299 10 ADE09872  
Ade09872 Novel DNA  
c 980 17.2 51.8 1322 6 ABK34963  
Abk34963 Human cdn  
c 981 17.2 51.8 1330 2 AAX06100  
Aax06100 Human dna  
c 982 17.2 51.8 1335 2 AAQ81700  
Aaq81700 C2P2(75-5  
c 983 17.2 51.8 1337 10 ADE79052  
Ade79052 Human pro  
c 984 17.2 51.8 1344 13 ADX52010  
Adx52010 Plant ful  
c 985 17.2 51.8 1380 4 AAK16799  
Aak16799 Human bra  
c 986 17.2 51.8 1380 4 ABS42183  
Abs42183 Human liv  
c 987 17.2 51.8 1380 5 AAI08958  
Aai08958 Probe #89  
c 988 17.2 51.8 1407 10 ADE79039  
Ade79039 Human pro  
c 989 17.2 51.8 1409 10 ADE79066  
Ade79066 Human pro  
c 990 17.2 51.8 1424 6 ABQ34623  
Abq34623 Oligonuc1  
c 991 17.2 51.8 1424 6 ABQ34622  
Abq34622 Oligonuc1  
c 992 17.2 51.8 1448 10 ADE79040  
Ade79040 Human pro  
c 993 17.2 51.8 1464 6 ABZ15412  
Abz15412 Arabidops  
c 994 17.2 51.8 1475 10 ADF81935  
Adf81935 Leukaemia  
c 995 17.2 51.8 1503 4 AAF32484  
Aaf32484 A. tumefa  
c 996 17.2 51.8 1549 4 ABL16991  
Abl16991 Drosophil  
c 997 17.2 51.8 1574 6 RAD30575  
Rad30575 Human pro  
c 998 17.2 51.8 1591 6 ACA47751  
Aca47751 Prokaryot  
c 999 17.2 51.8 1616 14 ADZ81525  
Adz81525 L. plant  
1000 17.2 51.8 1623 2 AAZ96344  
Aaz96344 S. pneumo

## ALIGNMENTS

RESULT 1  
ADI28530  
ID ADI28530 standard; DNA; 42 BP.  
XX AC ADI28530;  
XX DT 22-APR-2004 (first entry)  
XX TT Human GPCR retinoic acid induced 3 (RAI-3) peptide DNA.  
XX DE  
XX DE Human GPCR retinoic acid induced 3 (RAI-3) peptide DNA.  
XX XX Retinoic acid induced 3; RAI-3; human; G-protein coupled receptor; GPCR;  
XX KW antiinflammatory; immunosuppressive; cytostatic; cardiant; antiallergic;  
XX KW broncholytic; gene therapy; ss.  
XX XX Homo sapiens.  
XX OS Synthetic.  
XX XX WO2004001060-A2.  
XX XX 31-DEC-2003.  
XX XX 20-JUN-2003; 2003WO-US019255.  
XX XX 20-JUN-2002; 2002US-0390850P.  
XX PR 29-AUG-2002; 2002US-0407006P.  
XX XX (BRIM ) BRISTOL-MYERS SQUIBB CO.  
XX PA  
XX FI Whitney GS, Opitck G, Garulacan L, Ramanathan CS, McKinnon M;  
XX PI Bennett KL, Barber LE, Cacace A, Tsuchihashi Z;  
XX XX WPI; 2004-090973/09.  
XX XX New nucleic acid molecule encoding a human G-protein coupled receptor.  
XX PT RAI3, useful for preventing, treating or ameliorating chronic obstructive

PT pulmonary disease (COPD), COPD-like disorder, or the underlying symptoms  
of COPD.

XX Disclosure; SEQ ID NO 33; 301pp; English.

XX The present sequence is that of DNA encoding a peptide corresponding to  
amino acids 340-353 of a human G-protein coupled receptor, denoted  
retinoic acid induced 3 (RAI-3) protein ADI28460. The peptide was  
identified in studies of RAI-3 phosphorylation in response to cigarette  
smoke. Primers based on this sequence can be used in a PCR to identify  
individuals at risk for developing chronic obstructive pulmonary disease  
(COPD). RAI-3 is tyrosine phosphorylated and/or is associated/complexed  
with tyrosine phosphorylated proteins only in cells that have been  
exposed to cigarette smoke. Since RAI-3 is primarily expressed in lung  
tissue, and since cigarette smoke is a major causative factor of COPD,  
RAI-3 provides a novel cellular target for identifying modulators, e.g.  
agonists or antagonists, useful for the treatment and/or prevention of  
COPD and related disorders such as emphysema and chronic bronchitis. RAI-  
3 modulators, e.g. agonists and antagonists, especially antisense  
compounds, can be used to treat COPD and other disorders an diseases  
associated with regulation of NF-kB and/or its associated or interacting  
signaling molecules. Single nucleotide polymorphisms (SNPs) detected in  
the RAI-3 gene are useful for determining COPD association in  
individuals. RAI-3 nucleic acid molecules and polypeptides are useful for  
preventing, treating or ameliorating disorders related to aberrant GPCR  
signaling or cell cycle regulation, pulmonary disorders, inflammatory  
lung disorders, COPD, the underlying symptoms of COPD, COPD-related  
disorders or conditions, autoimmune disorders, disorders related to  
hyperimmune activity, inflammatory conditions, disorders related to  
aberrant acute phase responses, hypercongenital conditions, birth  
defects, necrotic lesions, wounds, organ transplant rejection, renal  
diseases, ischaemia-reperfusion injury, heart disorders, disorders  
related to aberrant signal transduction, proliferation disorders,  
cancers, HIV infection, asthma, cystic fibrosis, pulmonary fibrosis,  
ulcerative colitis, cerebral infarct, myocardial infarct, diabetic  
nephropathy, allergic rhinitis, Crohn's disease, atherosclerosis,  
rheumatoid arthritis, inflammatory/autoimmune disorders, glioblastoma,  
pulmonary small cell undifferentiated carcinoma, carcinoma of the breast,  
colon, lung, ovary, pancreas, prostate, non-Hodgkin's lymphoma, disorders  
associated with aberrant cell adhesion, I-CAM function and/or regulation,  
E-selectin function and/or regulation, or aberrant NF-kB function and/or  
regulation (all claimed).

XX SQ Sequence 42 BP; 9 A; 8 C; 7 G; 5 T; 0 U; 13 Other;

Query Match 100.0%; Score 33.2; DB 12; Length 42;

Best Local Similarity 100.0%; Pred. No. 4.2e-05;

Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCNCAYCNCNTGCCNTCCNTAYARGAYTAYGARGCTNAAR 42

Db 1 GCNCAYCNCNTGCCNTCCNTAYARGAYTAYGARGCTNAAR 42

## RESULT 2

ADT96418

ID ADT96418 standard; cDNA; 497 BP.

XX AC ADT96418;

XX DT 16-DEC-2004 (first entry)

XX DE Colon cancer associated human cDNA sequence #1925.

XX KW Colon cancer; T cell; tumour protein; C6345; C635S; C637S; C640S; C636S;  
XX KW humoral immune response; cellular immune response; cytostatic;  
XX KW immunostimulant; human; ss.

XX OS Homo sapiens.

XX XX US2003087818-A1.

XX XX 08-MAY-2003.



```
XX 01-FEB-2002; 2002US-00066543.
PF 02-FEB-2001; 2001US-0267400P.
XX 07-FEB-2001; 2001US-0267382P.
PR 11-MAY-2001; 2001US-0290322P.
PR 12-JUL-2001; 2001US-0305265P.
XX 16-AUG-2001; 2001US-0313077P.
XX (CORI-) CORIXA CORP.
XX Jiang Y, Chenault RA, Xu J, Indirias CY, Lodes MJ, Secrhist H;
XX Carter D, Fanger GR, Smith CL, Durham M, Stolk JA;
XX WPI; 2003-040540/03.
XX New isolated nucleic acids and polypeptides capable of eliciting a
XX humoral and/or cellular immune response, useful for diagnosing,
XX preventing or treating cancer, particularly colon cancer.
XX Claim 1; SEQ ID NO 1937; 87pp; English.
XX The invention relates to polynucleotide and polypeptide sequences
XX associated with cancer, particularly colon cancer. Also disclosed are (i)
XX an expression vector comprising the polynucleotide, (ii) a host cell
XX transformed or transfected with the expression vector, (iii) an isolated
XX antibody, or its antigen-binding fragment, which specifically binds to
XX the polypeptide, (iv) a method of detecting or determining the presence
XX of cancer in a patient, (v) a fusion protein comprising at least one of
XX the polypeptides, (vi) an oligonucleotide that hybridises to the
XX polynucleotide sequence under highly stringent conditions, and (vii) a
XX method of stimulating and/or expanding T cells specific for a tumour
XX protein. The polypeptide specifically comprises the amino acid sequence
XX of C634S, C635S, C637S, C640S, C636S or one of the potential open reading
XX frames (ORFs) of C636S. These polypeptides are encoded by the
XX polynucleotide sequences, where both are capable of eliciting a humoral
XX and/or cellular immune response. The polynucleotides, polypeptides, and
XX antibodies are useful for diagnosing, preventing or treating cancer,
XX particularly colon cancer. The polynucleotide and polypeptide sequences
XX are also useful in DNA strand invasion, antisense inhibition, mutational
XX analysis, nucleic acid purification, isolation of transcriptionally
XX active genes, blocking or transcription factor binding, genome cleavage
XX or in situ hybridisation, and as enhancers of transcription or
XX biomarkers. This sequence represents a human colon cancer associated
XX cDNA. Note: The sequence data for this patent was obtained in electronic
XX format directly from the USPTO web site at seqdata.uspto.gov
XX SQ Sequence 497 BP; 126 A; 130 C; 129 G; 112 T; 0 U; 0 Other;
Query Match 90.4%; Score 30; DB 11; Length 497;
Best Local Similarity 64.3%; Pred. No. 0.0025;
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;
Oy 1 GCNCAYGNTGCCNCCNTAARGAYTAYGARGTNAAR 42
Db 344 GCCCAGCTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 385
RESULT 3
ADX42900
XX ADX42900 standard; cDNA; 497 BP.
XX AC ADX42900;
XX XX
XX 21-APR-2005 (first entry)
XX DE Human cDNA encoding colon cancer protein SEQ ID NO 1937.
XX XX
XX Cytostatic; Immunostimulant; therapy; diagnosis; colon cancer; neoplasm;
XX ss; gene.
XX KW
XX OS Homo sapiens.
XX XX
```

```
PN WO200274156-A2.
XX 26-SEP-2002.
XX 01-FEB-2002; 2002WO-US002870.
XX 02-FEB-2001; 2001US-0267400P.
PR 07-FEB-2001; 2001US-0267382P.
XX 11-MAY-2001; 2001US-0290322P.
PR 12-JUL-2001; 2001US-0305265P.
XX 16-AUG-2001; 2001US-0313077P.
XX (CORI-) CORIXA CORP.
XX Jiang Y, Chenault RA, Xu J, Indirias CY, Lodes MJ, Secrhist H;
XX Carter D, Fanger GR, Smith CL, Durham M, Stolk JA;
XX WPI; 2003-040540/03.
XX New isolated nucleic acids and polypeptides capable of eliciting a
XX humoral and/or cellular immune response, useful for diagnosing,
XX preventing or treating cancer, particularly colon cancer.
XX Claim 1; SEQ ID NO 1937; 244pp; English.
XX The invention relates to a new isolated nucleic acid. The nucleic acids,
XX polypeptides, antibodies are useful for diagnosing, preventing or
XX treating cancer, particularly colon cancer. The nucleic acid and
XX polypeptides are also useful in DNA strand invasion, antisense
XX inhibition, mutational analysis, nucleic acid purification, isolation of
XX transcriptionally active genes, blocking or transcription factor binding,
XX genome cleavage or in situ hybridization, and as enhancers of
XX transcription or biomarkers. The kits are useful for detecting antibody
XX binding. The present sequence represents a human cDNA encoding a colon
XX cancer protein.
XX SQ Sequence 497 BP; 126 A; 130 C; 129 G; 112 T; 0 U; 0 Other;
Query Match 90.4%; Score 30; DB 11; Length 497;
Best Local Similarity 64.3%; Pred. No. 0.0025;
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;
Oy 1 GCNCAYGNTGCCNCCNTAARGAYTAYGARGTNAAR 42
Db 344 GCCCAGCTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 385
RESULT 4
ADT96292/c
XX ADT96292 standard; cDNA; 552 BP.
XX AC ADT96292;
XX XX
XX 16-DEC-2004 (first entry)
XX DT Colon cancer associated human cDNA sequence #1799.
XX DE
XX DE Colon cancer; T cell; tumour protein; C634S; C635S; C637S; C640S; C636S;
XX KW humoral immune response; cellular immune response; cytostatic;
XX KW immunostimulant; human; ss.
XX XX
XX OS Homo sapiens.
XX XX
XX US2003087818-A1.
XX PN
XX 08-MAY-2003.
XX PD
XX 01-FEB-2002; 2002US-00066543.
XX PF
XX 02-FEB-2001; 2001US-0267400P.
PR 07-FEB-2001; 2001US-0267382P.
XX 11-MAY-2001; 2001US-0290322P.
PR 12-JUL-2001; 2001US-0305265P.
XX 16-AUG-2001; 2001US-0313077P.
XX (CORI-) CORIXA CORP.
```

```
PR 16-AUG-2001; 2001US-0313077P.
XX (CORI-) CORIXA CORP.
PA
XX Jiang Y, Chenault RA, Xu J, Indirias CY, Lodes MJ, Secretist H;
XX Carter D, Fanger GR, Smith CL, Durham M, Stolk JA;
XX WPI; 2003-040540/03.
XX
XX New isolated nucleic acids and polypeptides capable of eliciting a
XX humoral and/or cellular immune response, useful for diagnosing,
XX preventing or treating cancer, particularly colon cancer.
XX
XX Claim 1; SEQ ID NO 1811; 87dp; English.
XX
XX The invention relates to polynucleotide and polypeptide sequences
XX associated with cancer, particularly colon cancer. Also disclosed are (i)
XX an expression vector comprising the polynucleotide, (ii) a host cell
XX transformed or transfected with the expression vector, (iii) an isolated
XX antibody, or its antigen-binding fragment, which specifically binds to
XX the polypeptide, (iv) a method of detecting or determining the presence
XX of cancer in a patient, (v) a fusion protein comprising at least one of
XX the polypeptides, (vi) an oligonucleotide that hybridizes to the
XX polynucleotide sequence under highly stringent conditions, and (vii) a
XX method of stimulating and/or expanding T cells specific for a tumour
XX protein. The polypeptide specifically comprises the amino acid sequence
XX of C634S, C635S, C637S, C638S, C636S or one of the potential open reading
XX frames (ORFs) of C636S. These polypeptides are encoded by the
XX polynucleotide sequences, where both are capable of eliciting a humoral
XX and/or cellular immune response. The polynucleotides, polypeptides, and
XX antibodies are useful for diagnosing, preventing or treating cancer,
XX particularly colon cancer. The polynucleotide and polypeptide sequences
XX are also useful in DNA strand invasion, antisense inhibition, mutational
XX analysis, nucleic acid purification, isolation of transcriptionally
XX active genes, blocking or transcription factor binding, genome cleavage
XX or in situ hybridization, and as enhancers of transcription or
XX biomarkers. This sequence represents a human colon cancer associated
XX cDNA. Note: The sequence data for this patent was obtained in electronic
XX format directly from the USPTO web site at seqdata.uspto.gov
XX
XX Sequence 552 BP; 126 A; 141 C; 148 G; 137 T; 0 U; 0 Other;
SQ
Query Match 90.4%; Score 30; DB 11; Length 552;
Best Local Similarity 64.3%; Pred. No. 0.0026;
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;
QY 1 GCNCAYCGTGGCCNTCNCNTAARGAYTAYGARGTNAAR 42
DB 154 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAGTAAAG 113
RESULT 5
ADX42774/C
ID AOX42774 standard; cDNA; 552 BP.
XX
XX ADX42774;
AC
XX
XX 21-APR-2005 (first entry)
DT
XX
XX Human cDNA encoding colon cancer protein SEQ ID NO 1811.
DE
XX
XX Cytostatic; Immunostimulant; therapy; diagnosis; colon cancer; neoplasm;
XX ss; gene.
XX
XX Homo sapiens.
OS
XX
XX WO200274156-A2.
EN
XX
XX 26-SEP-2002.
PD
XX
XX 01-FEB-2002; 2002WO-US002870.
XX
XX 02-FEB-2001; 2001US-0267400P.
XX
```

```
PR 07-FEB-2001; 2001US-0267382P.
PR 11-MAY-2001; 2001US-0290322P.
PR 12-JUL-2001; 2001US-0305265P.
PR 16-AUG-2001; 2001US-0313077P.
XX
XX (CORI-) CORIXA CORP.
PA
XX Jiang Y, Chenault RA, Xu J, Indirias CY, Lodes MJ, Secretist H;
XX Carter D, Fanger GR, Smith CL, Durham M, Stolk JA;
XX WPI; 2003-040540/03.
XX
XX New isolated nucleic acids and polypeptides capable of eliciting a
XX humoral and/or cellular immune response, useful for diagnosing,
XX preventing or treating cancer, particularly colon cancer.
XX
XX Claim 1; SEQ ID NO 1811; 244pp; English.
XX
XX The invention relates to a new isolated nucleic acid. The nucleic acids,
XX polypeptides, antibodies are useful for diagnosing, preventing or
XX treating cancer, particularly colon cancer. The nucleic acid and
XX polypeptides are also useful in DNA strand invasion, antisense
XX inhibition, mutational analysis, nucleic acid purification, isolation of
XX transcriptionally active genes, blocking or transcription factor binding,
XX genome cleavage or in situ hybridization, and as enhancers of
XX transcription or biomarkers. The kits are useful for detecting antibody
XX binding. The present sequence represents a human cDNA encoding a colon
XX cancer protein.
XX
XX Sequence 552 BP; 126 A; 141 C; 148 G; 137 T; 0 U; 0 Other;
SQ
Query Match 90.4%; Score 30; DB 11; Length 552;
Best Local Similarity 64.3%; Pred. No. 0.0026;
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;
QY 1 GCNCAYCGTGGCCNTCNCNTAARGAYTAYGARGTNAAR 42
DB 154 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAGTAAAG 113
RESULT 6
AAZ90046
ID AAZ90046 standard; cDNA; 603 BP.
XX
XX AAZ90046;
AC
XX
XX 09-MAY-2000 (first entry)
DT
XX
XX Hydrophobic domain containing protein clone HP10549 coding sequence.
DE
XX
XX Hydrophobic domain; clone HP10549; nutritional supplement; SCID; HIV;
XX cell proliferation; immune stimulant; immune deficiency; tumour; pain;
XX rheumatoid arthritis; insulin dependent diabetes mellitus; fertility;
XX myasthenia gravis; haematopoiesis regulator; tissue growth; depression;
XX anti-inflammatory; infection; bodily characteristic; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200000506-A2.
XX
XX 06-JAN-2000.
PD
XX
XX 18-JUN-1999; 99WO-JP003242.
XX
XX 26-JUN-1998; 98JP-00180008.
XX
XX (SAGA ) SAGAMI CHEM RES CENT.
XX (PROT-) PROTEGENE INC.
XX
XX Kato S, Kimura T;
XX
XX WPI; 2000-160665/14.
XX
XX P-PSDB; AAY78809.
XX
```

XX Novel human proteins having hydrophobic domains used for research and  
PT diagnostic purposes.  
XX  
XX  
PS Claim 3; Page 88; 117pp; English.  
XX  
CC This sequence represents the hydrophobic domain containing protein, clone  
CC HP10549 coding region. The sequence is isolated from a human stomach  
CC cancer cell line. The invention relates to human proteins with  
CC hydrophobic domains, the DNA and the cDNA encoding them. The  
CC polynucleotides and proteins are predicted to have biological activities  
CC which make them suitable for treating, preventing or ameliorating medical  
CC conditions in humans and animals. Suggested activities include  
CC nutritional activity (nutritional source or supplement); cytokine and  
CC cell proliferation/differentiation activity; immune stimulating (e.g. as  
CC vaccines) or suppressing activity (e.g. to treat various immune  
CC deficiencies such as AIDS or HIV, connective tissue disease, systemic  
CC lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary  
CC inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin  
CC dependent diabetes mellitus, myasthenia gravis, graft-versus-host disease  
CC and autoimmune inflammatory eye disease, as well as asthma, allergies and  
CC organ transplantation); haematopoiesis regulating activity (e.g. in  
CC treatment of myeloid or lymphoid cell deficiencies); tissue growth  
CC activity (e.g. wound healing and tissue repair, ulcers, burns,  
CC periodontal disease); activin/inhibin activity; chemotactic/chemokinetic  
CC activity; haemostatic and thrombolytic activity (e.g. treating  
CC haemophilias); receptor/ligand activity; anti-inflammatory activity; and  
CC tumour inhibition activity. The polynucleotides are also stated to be  
CC useful for gene therapy. Other activities include inhibiting infections  
CC caused by bacteria, fungi, viruses and other parasites (e.g. Hepatitis,  
CC malaria); effecting bodily characteristics such as, e.g. weight, colour,  
CC skin, effecting biorhythms or cardiac cycles; enhancing fertility;  
CC treatment of depression; treatment of pain; hormonal or endocrine  
CC activity. The polynucleotides may also be used for recombinant expression  
CC of the protein  
XX  
SQ Sequence 603 BP; 133 A; 179 C; 137 G; 154 T; 0 U; 0 Other;

Query Match 90.4%; Score 30; DB 3; Length 603;  
Best Local Similarity 64.3%; Pred. No. 0.0026;  
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

Qy 1 GCNCAYCNCNTGCCNTCCNTAYARGAYTAYGARGTNAAR 42  
|||:||||| |||:|||||:|||||:|||||:  
Db 550 GCCACGCTTGGCGCGCCCTTACAAAGACTATGAGTAAG 591

RESULT 7  
ABQ58527/c  
ID ABQ58527 standard; cDNA; 620 BP.  
AC ABQ58527;  
XX  
XX 02-AUG-2002 (first entry)  
DT  
DE Human colon cancer related nucleotide sequence SEQ ID NO:2222.  
XX  
XX Human; colon cancer; cancer; tissue profiling; forensic; mapping;  
KW genetic analysis; diagnostic; antisense therapy; gene; ss.  
XX  
XX Homo sapiens.  
OS  
XX WO200229086-A2.  
PN  
XX 11-APR-2002.  
PD  
XX 02-OCT-2001; 2001WO-US030732.  
PF  
XX 02-OCT-2000; 2000US-0237271P.  
PR  
XX (FARB ) BAYER CORP.  
XX  
PI Burgess C, Astle JH, Carroll E, Catino TJ, Dwivedi P, Molino GA;

PI Thiaglingam A, Lewis ME;  
XX  
XX WPI; 2002-426115/45.  
XX  
PT New isolated nucleic acid that is differentially expressed in cancer  
PT tissues useful for determining the presence of colon cancer in a cell or  
PT tissue type, and in antisense therapy.  
XX  
XX Claim 1; Fig 1; 796pp; English.  
XX  
CC ABQ56306 to ABQ60787 represent isolated nucleic acids (I) differentially  
CC expressed in cancer tissues. ABQ78993 to ABQ79004 represent proteins  
CC encoded by the ABQ60787 to ABQ60787 nucleic acid sequences. (I) can be  
CC used in antisense therapy. An antibody immunoreactive with a polypeptide  
CC encoded by (I) is useful for detecting cancer in a patient sample and  
CC for detecting the presence or absence of a polynucleotide encoded by a  
CC nucleic acid which hybridises to (I) in a cell. A probe/primer derived  
CC from (I) can be used for determining the presence of a nucleic acid which  
CC hybridises to (I), and for determining the phenotype of cells in a sample  
CC of cells from a patient. (I) is useful for determining the presence of  
CC colon cancer in a cell or tissue type, for determining the presence or  
CC state of other type of cancer, in antisense therapy, to generate  
CC macroarrays on a solid surface, to identify a chromosome on which the  
CC corresponding gene resides, and in tissue profiling, forensics, genetic  
CC analysis, mapping and diagnostic applications. (I) can be used to raise  
CC antibodies, and to screen for peptide analogues and antagonists  
XX  
SQ Sequence 620 BP; 152 A; 154 C; 161 G; 142 T; 0 U; 11 Other;  
Query Match 90.4%; Score 30; DB 6; Length 620;  
Best Local Similarity 64.3%; Pred. No. 0.0027;  
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;  
Qy 1 GCNCAYCNCNTGCCNTCCNTAYARGAYTAYGARGTNAAR 42  
|||:||||| |||:|||||:|||||:|||||:  
Db 154 GCCACGCTTGGCGCGCCCTTACAAAGACTATGAGTAAG 113

RESULT 8  
ABQ59698/c  
ID ABQ59698 standard; cDNA; 634 BP.  
XX  
XX ABQ59698;  
XX  
XX 02-AUG-2002 (first entry)  
DT  
DE Human colon cancer related nucleotide sequence SEQ ID NO:3393.  
XX  
XX Human; colon cancer; cancer; tissue profiling; forensic; mapping;  
KW genetic analysis; diagnostic; antisense therapy; gene; ss.  
XX  
XX Homo sapiens.  
OS  
XX WO200229086-A2.  
PN  
XX 11-APR-2002.  
PD  
XX 02-OCT-2001; 2001WO-US030732.  
PF  
XX 02-OCT-2000; 2000US-0237271P.  
PR  
XX (FARB ) BAYER CORP.  
PA  
XX Burgess C, Astle JH, Carroll E, Catino TJ, Dwivedi P, Molino GA;  
PI Thiaglingam A, Lewis ME;  
XX  
XX WPI; 2002-426115/45.  
XX  
XX New isolated nucleic acid that is differentially expressed in cancer  
XX tissues useful for determining the presence of colon cancer in a cell or  
XX tissue type, and in antisense therapy.  
XX  
XX Claim 1; Fig 1; 796pp; English.

XX ABQ56306 to ABQ60787 represent isolated nucleic acids (I) differentially  
CC expressed in cancer tissues. ABB78993 to ABB79004 represent proteins  
CC encoded by the ABQ60776 to ABQ60787 nucleic acid sequences. (I) can be  
CC used in antisense therapy. An antibody immunoreactive with a polypeptide  
CC encoded by (I) is useful for detecting cancer in a patient sample, and  
CC for detecting the presence or absence of a polynucleotide encoded by a  
CC nucleic acid which hybridizes to (I) in a cell. A probe/primer derived  
CC from (I) can be used for determining the presence of a nucleic acid which  
CC hybridizes to (I), and for determining the phenotype of cells in a sample  
CC of cells from a patient. (I) is useful for determining the presence of  
CC colon cancer in a cell or tissue type, for determining the presence or  
CC state of other type of cancer, in antisense therapy, to generate  
CC macroarrays on a solid surface, to identify a chromosome on which the  
CC corresponding gene resides, and in tissue profiling, forensics, genetic  
CC analysis, mapping and diagnostic applications. (I) can be used to raise  
CC antibodies, and to screen for peptide analogues and antagonists  
XX Sequence 634 BP: 153 A; 154 C; 168 G; 146 T; 0 U; 13 Others:

Query Match	90.4%	Score 30	DB 6	Length 634
Best Local Similarity	64.3%	Pred. No. 0.0027		
Matches 27	Conservative 7	Mismatches 8	Indels 0	Gaps 0
Qy	1	GCNCAYCNCITGGCCNTCCNCTAYAAARGAYTAYGARGTNAAR	42	
Db	158	GCCACACCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG	117	

RESULT 9	
ABT22328/c	
ID	ABT22328 standard; DNA; 642 BP.
XX	
XX	
AC	ABT22328;
XX	
DT	16-APR-2003 (first entry)
XX	
DE	Breast cancer marker gene SEQ ID No 701.
XX	
KW	Cytostatic; vaccine; breast cancer marker gene; breast mass; immunogen;
KW	chemotherapy; tumour burden; bait protein; two-hybrid; three-hybrid;
KW	surrogate marker gene; pharmacodynamic marker gene; transgenic animal;
KW	human; ds.
XX	
OS	Homo sapiens.
XX	
PN	WO200285298-A2.
XX	
PD	31-OCT-2002.
XX	
PF	19-APR-2002; 2002WO-US012612.
XX	
PR	20-APR-2001; 2001US-0285163P.
XX	
PA	(MILL-) MILLENNIUM PHARM INC.
XX	
PI	Lillie J, Palermo A, Wang Y, Steinmann K, Elias J, Mertens M;
XX	
DR	WPI; 2003-093053/08.

useful for evaluating a patient before, after or during therapy, to evaluate the reduction in a tumour burden. The breast cancer marker gene proteins are useful as immunogens for raising antibodies, by immunising a mammal with a breast cancer marker protein. The marker proteins are useful as bait proteins in a two-hybrid or three-hybrid assay, to identify other proteins which bind to or interact with the marker proteins. The breast cancer marker genes are useful as surrogate marker genes for one or more disorders, disease states or conditions leading to disease states, in particular, breast cancers. The breast cancer marker genes are useful as pharmacodynamic marker genes. An antibody which selectively binds to a protein of a breast cancer marker gene is useful for treating cancers, particularly breast cancers. The host cell of the invention is useful for producing non-human transgenic animals. This polynucleotide sequence represents one of the breast cancer marker genes of the invention

Sequence 642 BP: 148 A; 160 C; 180 G; 142 T; 0 U; 12 Other  
xx

```

Query Match          90.4%; Score 30; DB 10; Length 642;
Best Local Similarity 64.3%; Pred. No. 0.0021;
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

Qy      1 GCNCAYGCNTGGCCNCTCNCNTYAAARGAVTAYGAGRTNAAR   42
         |||:||||| ||||| ||||| :|||:|||||:|||||:
Db     184 GCCCACGCTTGCGCGAGCCCTTACAAAGACTATGAAGTAAG   143

RESULT 10
ADO30035
ID ADO30035 standard; cDNA; 1074 BP.
XX AC
XX AC ADO30035;
XX XX
DT DT 29-JUL-2004 (first entry)
XX XX
DE Human GPCR RAI3 polynucleotide. SEQ ID NO:1137.
```







```
DR WPI; 2005-371577/38.
DR P-PSDB; AEA00710.
XX
PT Novel isolated antibody e.g. anti-E16 or anti-TAT112 antibody that binds
PT to tumor-associated antigenic target polypeptide, useful for diagnosing
PT or treating cancer.
XX
XX Claim 1; SEQ ID NO 64; 96pp; English.
XX
CC This invention relates to a novel isolated antibody, for example anti-E16
CC or anti-TAT112 antibody, that binds to a tumor-associated antigenic
CC target polypeptide (TAT) and that lacks an associated signal peptide
CC sequence. The invention may be useful for the development of compounds
CC with a cytostatic activity acting as antagonists of the TAT188
CC polypeptide or RNA interference whilst the disclosed sequences may be
CC useful for gene therapy. The invention is useful for inducing the death
CC of a cell (such as a cancer cell chosen from breast, colon, rectum,
CC endometrium, kidney, lung, ovary, skin and liver) to which it binds,
CC inhibiting proliferation or promoting cell death of a cell expressing
CC TAT188. In addition, the invention may be useful for detecting the level
CC of TAT188 polypeptide in a test cell relative to a control cell, or
CC treating or preventing a cell proliferative disorder associated with
CC increased expression of TAT188. The novel antibody of the invention is
CC useful for inhibiting the growth of a cancer cell and may be useful for
CC diagnosing or treating cancer. The present sequence is that of the human
CC TAT143 cDNA which encodes a protein against which an antibody of the
CC invention may be targeted.
XX
XX Sequence 1460 BP; 302 A; 418 C; 362 G; 378 T; 0 U; 0 Other;
SQ
Query Match 90.4%; Score 30; DB 14; Length 1460;
Best Local Similarity 64.3%; Pred. No. 0.0034;
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;
Qy 1 GCNCAYGCGTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1186
Db 1145 GCCCAGCGTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1186
RESULT 17
AAF58615
ID AAF58615 standard; cDNA; 1619 BP.
XX
AC AAF58615;
XX
DT 24-APR-2001 (first entry)
XX
DE Human RECAP polynucleotide, SEQ ID NO: 43.
XX
KW Human; RECAP; receptors and associated proteins; cerebroprotective;
KW nootropic; neuroprotective; anticonvulsant; antiparkinsonian; anti-HIV;
KW antidiabetic; immunostimulant; immunomodulator; antiinflammatory;
KW antithyroid; immunosuppressive; nephrotropic; antigout; thyromimetic;
KW cytostatic; antibacterial; virucide; fungicide; protozoacide;
KW antiarteriosclerotic; hepatotropic; gene therapy; infection; cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200107612-A2.
XX
PD 01-FEB-2001.
XX
PF 21-JUL-2000; 2000WO-US020035.
XX
PR 21-JUL-1999; 95US-0145232P.
PR 07-OCT-1999; 95US-0158578P.
PR 12-NOV-1999; 99US-0165192P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Au-Young J, Bandman O, Tang YT, Yue H, Azimzai Y, Burford N;
PI Baughn MR, Lu DM, Hillman JL, Patterson C, Lal P;
XX
```

```
DR WPI; 2001-168554/17.
DR P-PSDB; AAB68891.
XX
PT Novel receptors and associated proteins for diagnosis and treatment of
PT neurological disorders, immunological disorders including autoimmune/
PT inflammatory disorders and cell proliferative disorders such as cancer.
XX
XX Example 5; Page 126-127; 128pp; English.
XX
CC The present sequence encodes a human RECAP (receptors and associated
CC proteins) polypeptide. RECAP polynucleotides and polypeptides are useful
CC in the diagnosis, treatment and prevention of neurological disorders such
CC as stroke, Alzheimer's disease, Pick's disease, Huntington's disease,
CC dementia, Parkinson's disease, Down's syndrome, amyotrophic lateral
CC sclerosis, multiple sclerosis, bacterial and viral meningitis, CJD
CC (Creutzfeldt-Jakob disease), GSS (Gerstmann -Straussler-Scheinker
CC syndrome); immunological disorders, including autoimmune/inflammatory
CC disorders such as AIDS, Digeorge's syndrome, severe combined
CC immunodeficiency disease (SCID), Chediak-Higashi syndrome, Cushing's
CC disease, Addison's disease, autoimmune thyroiditis, Crohn's disease,
CC diabetes mellitus, Good pasture's syndrome, gout, Grave's disease,
CC Hashimoto's thyroiditis, Sjogren's syndrome, Werner's syndrome, viral,
CC bacterial, fungal, parasitic, protozoal, and helminthic infections; and
CC cell proliferation disorders such as arteriosclerosis, atherosclerosis,
CC cirrhosis, hepatitis and cancer
XX
XX Sequence 1619 BP; 331 A; 463 C; 400 G; 425 T; 0 U; 0 Other;
SQ
Query Match 90.4%; Score 30; DB 4; Length 1619;
Best Local Similarity 64.3%; Pred. No. 0.0035;
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;
Qy 1 GCNCAYGCGTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1178
Db 1137 GCCCAGCGTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1178
RESULT 18
AAZ90056
ID AAZ90056 standard; cDNA; 1718 BP.
XX
AC AAZ90056;
XX
DT 09-MAY-2000 (first entry)
XX
DE Hydrophobic domain containing protein clone HP10549 nucleotide sequence.
XX
KW Hydrophobic domain; clone HP10549; nutritional supplement; SCID; HIV;
KW cell proliferation; immune stimulant; immune deficiency; tumour; pain;
KW rheumatoid arthritis; insulin dependent diabetes mellitus; fertility;
KW myasthenia gravis; haematopoiesis regulator; tissue growth; depression;
KW anti-inflammatory; infection; bodily characteristic; ss.
XX
OS Homo sapiens.
XX
PN WO200000506-A2.
XX
PD 06-JAN-2000.
XX
PF 18-JUN-1999; 99WO-JP003242.
XX
PR 26-JUN-1998; 98JP-00180008.
XX
PA (SAGA ) SAGAMI CHEM RES CENT.
PA (PROT-) PROTEGENE INC.
XX
PI Kato S, Kimura T;
XX
XX WPI; 2000-160665/14.
XX P-PSDB; AAY78809.
XX
PT Novel human proteins having hydrophobic domains used for research and
PT diagnostic purposes.
```



XX PS Claim 4; Page 111-113; 117pp; English.

XX CC This sequence represents the hydrophobic domain containing protein, clone

CC HPI0549 nucleotide sequence. The sequence is isolated from a human

CC stomach cancer cell line. The invention relates to human proteins with

CC hydrophobic domains, the DNA and the cDNA encoding them. The

CC polynucleotides and proteins are predicted to have biological activities

CC which make them suitable for treating, preventing or ameliorating medical

CC conditions in humans and animals. Suggested activities include

CC nutritional activity (nutritional source or supplement); cytokine and

CC cell proliferation/differentiation activity; immune stimulating (e.g. as

CC vaccines) or suppressing activity (e.g. to treat various immune

CC deficiencies such as AIDS or HIV, connective tissue disease, systemic

CC lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary

CC inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin

CC dependent diabetes mellitus, myasthenia gravis, graft-versus-host disease

CC and autoimmune inflammatory eye disease, as well as asthma, allergies and

CC organ transplantation); haematopoiesis regulating activity (e.g. in

CC treatment of myeloid or lymphoid cell deficiencies); tissue growth

CC activity (e.g. wound healing and tissue repair, ulcers, burns,

CC periodontal disease); activin/inhibin activity; chemotactic/chemokinetic

CC activity; haemostatic and thrombolytic activity (e.g. treating

CC haemophilias); receptor/ligand activity; anti-inflammatory activity; and

CC tumour inhibition activity. The polynucleotides are also stated to be

CC useful for gene therapy. Other activities include inhibiting infections

CC caused by bacteria, fungi, viruses and other parasites (e.g. Hepatitis,

CC malaria); affecting bodily characteristics such as, e.g. weight, colour,

CC skin, affecting biorhythms or circadian cycles; enhancing fertility;

CC treatment of depression; treatment of pain; hormonal or endocrine

CC activity. The polynucleotides may also be used for recombinant expression

CC of the protein

XX CC

XX SQ Sequence 1718 BP; 381 A; 495 C; 377 G; 465 T; 0 U; 0 Other;

Query Match 90.4%; Score 30; DB 3; Length 1718;

Best Local Similarity 64.3%; Pred. No. 0.0035;

Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

Oy 1 GCNCAYCGTGGCCNTCNCNTAARGAYTAYGARGTNAAR 42

Db 561 GCCCAGCGTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 602

RESULT 19

ADF70573

ID ADF70573 standard; DNA; 1788 BP.

XX AC

XX ADP70573;

XX DT

XX DT 12-FEB-2004 (first entry)

XX DE Orphan receptor ligand-related human protein gene SeqID196.

XX KW ligand; orphan receptor protein; fusion protein; fluorescent protein;

KW cell expression; green fluorescent protein; GFP; GFP-1; wild-type GFP;

KW GFPuv; Enhanced GFP; EGFP; human; gene; ds.

XX OS Homo sapiens.

XX PN WO2003071272-A1.

XX PD

XX PD 28-AUG-2003.

XX PF 21-FEB-2003; 2003WO-JP001901.

XX PR 22-FEB-2002; 2002JP-00045728.

PR 23-JUL-2002; 2002JP-00213949.

PR 11-OCT-2002; 2002JP-00298237.

XX PA (TAKE ) TAKEDA CHEM IND LTD.

XX PI Hinuma S, Fujii R, Ogi K, Komatsu H, Kawamata Y, Hosoya M;

XX WPI; 2003-697654/66.

DR P-PSDB; ADF70471.

XX PT Transformation of cells with a fusion protein of an orphan receptor

PT protein with a fluorescent protein useful for identification of ligands

PT to the orphan receptor.

XX PS Example 4; SEQ ID NO 196; 594pp; Japanese.

XX CC This invention relates to a novel method of identifying ligands to an

CC orphan receptor protein which comprises transforming cells with DNA

CC encoding a fusion protein of the orphan receptor with a fluorescent

CC protein, so that the fusion protein is expressed in the cells (or cell

CC membranes isolated from them) and contacting the cells with the potential

CC ligand to be tested. A suitable fluorescent protein for incorporation in

CC the fusion protein is green fluorescent protein (GFP), for example GFP-1,

CC wild-type GFP, GFPuv or Enhanced GFP (EGFP). The method is useful for the

CC identification of ligands binding to an orphan receptor protein.

XX SQ Sequence 1788 BP; 444 A; 473 C; 412 G; 459 T; 0 U; 0 Other;

Query Match 90.4%; Score 30; DB 10; Length 1788;

Best Local Similarity 64.3%; Pred. No. 0.0036;

Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

Oy 1 GCNCAYCGTGGCCNTCNCNTAARGAYTAYGARGTNAAR 42

Db 1018 GCCCAGCGTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1059

RESULT 20

ADF10173

ID ADF10173 standard; cDNA; 2302 BP.

XX AC

XX ABT10173;

XX DT

DT 04-DEC-2002 (first entry)

XX DE Human breast cancer associated coding sequence SEQ ID NO: 307.

XX KW Human; breast specific gene; breast cancer; differential expression;

KW cytostatic; gene therapy; gene; ss.

XX OS Homo sapiens.

XX PN WO200259271-A2.

XX PD

XX PD 01-AUG-2002.

XX PF 25-JAN-2002; 2002WO-US002176.

XX PR 25-JAN-2001; 2001US-0263757P.

PR 25-APR-2001; 2001US-0286090P.

PR 23-MAY-2001; 2001US-0292517P.

XX PA (GENE-) GENE LOGIC INC.

XX PI Orr MS, Nation M, Diggans JC, Zeng W;

XX WPI; 2002-674803/72.

XX PT Diagnosing breast cancer in a patient comprises detecting the level of

PT gene expression in cell or tissue samples, where a differential gene

PT expression is indicative of breast cancer.

XX PS Claim 1; SEQ ID NO 307; 260pp + Sequence Listing; English.

XX CC The present invention relates to methods of diagnosing breast cancer in a

CC patient, which comprise detecting the level of expression in a tissue

CC sample of two or more genes selected from those shown in AB09867-

CC ABT11112, where a differential expression of the genes indicates breast

CC cancer. The methods are useful in diagnosing, treating, detecting the

CC	with up- or down-regulation of GPCR-like RAIGI activity, such as obesity,
CC	anorexia, cachexia or diabetes
XX	
SQ	Sequence 2302 BP; 494 A; 666 C; 533 G; 609 T; 0 U; 0 Other;
	Query Match 90.4%; Score 30; DB 8; Length 2302;
	Best Local Similarity 64.3%; Pred. No. 0.0038;
	Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;
Qy	1 GCNCAYGCNTGGCCGTCNCNTAYARGAYTAGTGTTNAAR 42
Db	1117 GCCACGCTGGCGGAGCCCTTACAAGACTGAAGATAAG 1158

```

RESULT 22
ADD93240
ID   ADD93240 standard; cDNA; 2302 BP.
XX
AC   ADD93240;
XX

```

DE	RAIG1 coding sequence.	
XX		
XX	ss; gene; retinoic acid-inducible gene 1 ; RAIG1; orphan ;	
KW	G-protein coupled receptor; GPCR; chromosome 12; GPCR5B; carcinoma ;	
KW	vaccine; breast cancer; pancreatic cancer; lung cancer; liver cancer;	
KW	ovarian cancer; colon cancer ; osteosarcoma.	
XX		
OS	Homo sapiens.	
XX		
XX		
FH	Key	Location/Qualifiers
FT	CDS	100..1173
FT		/*tag= a
FT		/product= "RAIG1"
XX		
XX		
PN	W02003087832-A2.	

23-OCT-2003.  
10-APR-2003; 2003WO-GB001587.  
11-APR-2002; 2002GB-00008331.  
17-SEP-2002; 2002GB-00021538.  
(OXFO-) OXFORD GLYCOSCIENCES UK LTD.  
Terrett JA;

WPI: 2003-845382/78.  
P-PSDE/ ADD93239.  
GENBANK/ AF095448.  
Screening, diagnosing and/or treating carcinoma, including breast, pancreatic, lung, liver, ovarian and colon cancer by detecting the change in expression or activity of an RAIG1 polypeptide or encoding nucleic acid molecule.

CC This sequence encodes retinoic acid-inducible gene 1 (RAIG1) polypeptide.  
 CC RAIG1 is an orphan G-protein coupled receptor (GPCR) located on  
 CC chromosome 12. RAIG1 shows a restricted expression pattern compared to a  
 CC related receptor, GPCR5B, which is widely expressed in peripheral and  
 CC

the method of the invention for screening for and/or diagnosis of carcinoma in a subject, and/or monitoring the effectiveness of carcinoma therapy. The method comprises detecting and/or quantifying in a biological sample obtained from the subject an RAIG1 polypeptide and a nucleic acid molecule. The RAIG1 polypeptide and nucleic acid molecule are useful in the manufacture of a medicament for the treatment of carcinoma, where the composition is a vaccine. An agent which interacts with or causes change in the expression or activity of an RAIG1

CC polypeptide or nucleic acid molecule, is also useful in the manufacture  
CC of a medicament for the treatment of carcinoma that is breast cancer,  
CC pancreatic cancer, lung cancer, liver cancer, ovarian cancer, colon  
CC cancer and/or osteosarcoma. They can also be used in the diagnosis and  
CC screening of such carcinomas.

XX Sequence 2302 BP; 494 A; 666 C; 533 G; 609 T; 0 U; 0 Other;

Query Match 90.4%; Score 30; DB 10; Length 2302;

Best Local Similarity 64.3%; Pred. No. 0.0038;

Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

Qy 1 GCNCAYCGTGGCCNTCCNTAYAAAGAYTAYGARGTNAAR 42

|||:||||| |||:||||| |||:||||| |||:||||| |||:|||||

Db 1117 GCCCAGCGTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1158

#### RESULT 23

ADL24773

ID ADL24773 standard; DNA; 2302 BP.

XX AC ADL24773;

XX 20-MAY-2004 (first entry)

XX Intestinal epithelium/peyer's patch M cell-associated DNA sequence #171.

XX intestinal epithelium cell development; peyer's patch M cell development;  
XX inflammatory bowel disease; glutenenteropathy; infectious disease;  
XX autoimmune disease; haemolytic anaemia; rheumatoid arthritis; dermatitis;  
XX Grave's disease; multiple sclerosis; allergy; asthma; diabetic mellitus;  
XX immune system disorder; hypersensitivity; anaphylaxis;  
XX blood group incompatibility; ds; human.

XX Homo sapiens.

XX WO200280852-A2.

XX 17-OCT-2002.

XX 04-APR-2002; 2002WO-US010873.

XX 04-APR-2001; 2001US-0281416P.

XX (DIGI-) DIGITAL GENE TECHNOLOGIES INC.

XX Brayden DJ, Byrne D, O'mahony DJ, Evans CF, Mah SP, Lo DD;

XX WPI; 2003-075470/07.

XX Novel isolated or purified polypeptide encoded by genes associated with  
XX intestinal epithelium or M cell development, differentiation or function,  
XX useful for treating autoimmune diseases and infectious diseases.

XX Claim 1; SEQ ID NO 283; 152pp; English.

XX The invention comprises DNA sequences which are associated with  
XX intestinal epithelium and peyer's patch M cells. The DNA sequences of the  
XX invention are useful for assessing, modifying, modulating or regulating  
XX intestinal epithelium or M cell development. The DNA sequences of the  
XX invention are also useful in the treatment of: inflammatory bowel  
XX disease, glutenenteropathy, infectious diseases, autoimmune diseases  
XX (e.g. haemolytic anaemia, rheumatoid arthritis, dermatitis, Grave's  
XX disease, multiple sclerosis, allergy, asthma and diabetic mellitus),  
XX diseases or disorders of the immune system, hypersensitivity,  
XX anaphylaxis, and blood group incompatibility. The present nucleic acid  
XX represents an intestinal epithelium/peyer's patch M cell-associated DNA  
XX sequence of the invention. NOTE: The present sequence is not shown in the  
XX specification, but has been retrieved from the WIPO website.

XX Sequence 2302 BP; 494 A; 666 C; 533 G; 609 T; 0 U; 0 Other;

Query Match 90.4%; Score 30; DB 10; Length 2302;

Best Local Similarity 64.3%; Pred. No. 0.0038;

Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

Qy 1 GCNCAYCGTGGCCNTCCNTAYAAAGAYTAYGARGTNAAR 42

|||:||||| |||:||||| |||:||||| |||:||||| |||:|||||

Db 1117 GCCCAGCGTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1158

#### RESULT 24

ABT31923

ID ABT31923 standard; DNA; 2316 BP.

XX AC ABT31923;

XX 01-MAY-2003 (first entry)

XX Human breast cancer / ovarian cancer related coding sequence #30.

XX Human; gene; ds; cytostatic; breast cancer; ovarian cancer.

XX Homo sapiens.

XX WO2003000012-A2.

XX 03-JAN-2003.

XX 21-JUN-2002; 2002WO-US019773.

XX 21-JUN-2001; 2001US-0300159P.

XX 27-JUN-2001; 2001US-0301351P.

XX (MILL-) MILLENNIUM PHARM INC.

XX Veiby OP;

XX WPI; 2003-267848/26.

XX P-PSDB; ABJ37054.

XX Determining the presence of breast cancer in an individual, involves  
XX using specific polynucleotide markers.

XX Disclosure; Page 163; 233pp; English.

XX The invention comprises a method for assessing whether a patient is  
XX afflicted with breast cancer or ovarian cancer. The method involves the  
XX use of specific DNA markers. The method of the invention is useful in the  
XX detection and treatment of ovarian and breast cancer. DNA sequences  
XX ABT31894 - ABT31949 encode human breast/ovarian cancer-related proteins

XX Sequence 2316 BP; 502 A; 670 C; 535 G; 609 T; 0 U; 0 Other;

Query Match 90.4%; Score 30; DB 10; Length 2316;

Best Local Similarity 64.3%; Pred. No. 0.0039;

Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

Qy 1 GCNCAYCGTGGCCNTCCNTAYAAAGAYTAYGARGTNAAR 42

|||:||||| |||:||||| |||:||||| |||:||||| |||:|||||

Db 1123 GCCCAGCGTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1164

#### RESULT 25

AAH14688

ID AAH14688 standard; cDNA; 2446 BP.

XX AC AAH14688;

XX 26-JUN-2001 (first entry)

XX Human cDNA sequence SEQ ID NO:12388.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

```

XX EP1074617-A2.
PN
XX
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-00116126.
XX
XX 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT length cDNAs defined in the specification, and for the detection and/or
PT diagnosis of the abnormality of the proteins encoded by the full-length
PT cDNAs.
XX
XX Claim 8; SEQ ID NO 12388; 2537pp + Sequence Listing; English.
XX
XX The present invention describes primer sets for synthesizing 5602 full-
CC length cDNAs defined in the specification. Where a primer set comprises:
CC (a) an oligo-dT primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention
XX
XX Sequence 2446 BP; 507 A; 723 C; 572 G; 644 T; 0 U; 0 Other;
SQ
Query Match 90.4%; Score 30; DB 4; Length 2446;
Best Local Similarity 64.3%; Pred. No. 0.0039;
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 GCNCAYCGNTGGCCNTCCNTAYARGAYTAYGAGCTNAAR 42
|||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1271 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1312

RESULT 26
ABZ42832
ID ABZ42832 standard; DNA; 2456 BP.
XX
XX ABZ42832;
XX
XX 04-MAR-2003 (first entry)
XX
XX Human G protein-coupled receptor RAIG1 nucleotide SEQ ID NO:453.
XX
XX G protein-coupled receptor; GPCR; antigenic peptide; gene therapy;
XX G protein-coupled receptor modulator; antibody; immune-related disease;

```

---

```

KW growth-related disease; cell regeneration-related disease; AIDS; cancer;
KW immunological-related cell proliferative disease; autoimmune disease;
KW Alzheimer's disease; atherosclerosis; infection; osteoarthritis; allergy;
KW osteoporosis; cardiomyopathy; inflammation; Crohn's disease; diabetes;
KW graft versus host disease; Parkinson's disease; multiple sclerosis; pain;
KW psoriasis; anxiety; depression; schizophrenia; dementia; memory loss;
KW mental retardation; epilepsy; asthma; tuberculosis; obesity; nausea;
KW hypertension; hypotension; renal disorder; rheumatoid arthritis; trauma;
KW ulcer; gene; ds.
XX
XX Homo sapiens.
OS
XX WO200261087-A2.
PN
XX 08-AUG-2002.
PD
XX 19-DEC-2001; 2001WO-US050107.
PF
XX 19-DEC-2000; 2000US-0257144P.
PR
XX (LIFE-) LIFESPAN BIOSCIENCES INC.
PA
XX Burmer GC, Roush CL, Brown JP;
PI
XX WPI; 2003-046718/04.
PI
XX P-PSDB; ABP81984.
DR
XX
XX New isolated antigenic peptides e.g., for G protein-coupled receptors
PT (GPCR), useful for diagnosing and designing drugs for treating conditions
PT in which GPCRs are involved, e.g. AIDS, Alzheimer's disease, cancer or
PT autoimmune diseases.
XX
XX Disclosure; Fig 1; 523pp; English.
XX
XX The present invention describes antigenic peptides (I) comprising: (a)
CC any one of 1601 sequences (see ABP82019 to ABP83619) of 12-24 amino
CC acids. Also described: (1) an assay for the detection of a particular G
CC protein-coupled receptor (GPCR) or a candidate polypeptide in a sample; or
CC and (2) an isolated antibody having high specificity and high affinity or
CC avidity for a particular GPCR. (I) can be used as GPCR modulators and in
CC gene therapy. The antigenic peptides for GPCRs are useful in detecting an
CC antibody against a particular GPCR, and in the production of specific
CC antibodies. The peptides and antibodies are also useful for detecting the
CC presence or absence of corresponding GPCRs. The antigenic peptides for
CC GPCRs and antibodies are useful for diagnosing and designing drugs for
CC treating immune-related diseases, growth-related diseases, cell
CC regeneration-related disease, immunological-related cell proliferative
CC diseases, or autoimmune diseases, e.g. AIDS, Alzheimer's disease,
CC atherosclerosis, bacterial, fungal, protozoan or viral infections,
CC osteoarthritis, osteoporosis, cancer, cardiomyopathy, chronic and acute
CC inflammation, allergies, Crohn's disease, diabetes, graft versus host
CC disease, Parkinson's disease, multiple sclerosis, pain, psoriasis,
CC anxiety, depression, schizophrenia, dementia, mental retardation, memory
CC loss, epilepsy, asthma, tuberculosis, obesity, nausea, hypertension, or
CC hypotension, renal disorders, rheumatoid arthritis, trauma, ulcers, or
CC any other disorder in which GPCRs are involved. The antibodies may be
CC used in immunoassays and immunodiagnosis. ABZ42523 to ABZ42869 encode
CC GPCR proteins given in ABP81675 to ABP82018, which are used in the
CC exemplification of the present invention
XX
XX Sequence 2456 BP; 519 A; 722 C; 571 G; 644 T; 0 U; 0 Other;
SQ
Query Match 90.4%; Score 30; DB 8; Length 2456;
Best Local Similarity 64.3%; Pred. No. 0.0039;
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 GCNCAYCGNTGGCCNTCCNTAYARGAYTAYGAGCTNAAR 42
|||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1271 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1312

RESULT 27
ACCT2695

```



CC also be useful in wound healing and in contraception. The present  
CC sequence represents a nucleic acid sequence of the invention.

XX  
SQ Sequence 2456 BP; 519 A; 722 C; 571 G; 644 T; 0 U; 0 Other;  
Query Match 90.4%; Score 30; DB 11; Length 2456;  
Best Local Similarity 64.3%; Pred. No. 0.0039;  
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 GCNCAYCGNTGGCCNTCCNTAYARGAYTAYGAGTNAAR 42  
DB 1271 GCCCAGCGTTGGCGGCCCTTACAAAGACTATGAAGTAAAG 1312

RESULT 29  
ADI28525  
ID ADI28525 standard; cDNA; 2456 BP.

XX AC ADI28525;

XX 22-APR-2004 (first entry)

XX Human GPCR retinoic acid induced 3 (RAI-3) cDNA.

XX Retinoic acid induced 3; RAI-3; human; G-protein coupled receptor; GPCR;  
KW antiinflammatory; immunosuppressive; cytostatic; cardiant; antiallergic;  
KW broncholytic; gene therapy; gene; single nucleotide polymorphism; SNP;  
KW chromosome 12p13-p12.3; ss.

XX Homo sapiens.

XX Key Location/Qualifiers  
FH variation replace(112,r)

FT /\*tag= a

FT /label= RAI-3-s1

FT /notes "located in 5' untranslated region"

FT /standard\_name= "Single nucleotide polymorphism"

FT 254..1327

FT /\*tag= b

FT /product= "Human RAI3"

FT replace(364,y)

FT /\*tag= c

FT /label= RAI-3-s2

FT /notes "exon 1, silent (Ala/Ala)"

FT /standard\_name= "Single nucleotide polymorphism"

FT replace(511,y)

FT /\*tag= d

FT /label= RAI-3-s3

FT /notes "exon 2, silent (Ile/Ile)"

FT /standard\_name= "Single nucleotide polymorphism"

FT replace(523,y)

FT /\*tag= e

FT /label= RAI-3-s4

FT /notes "exon 2, silent (Asp/Asp)"

FT /standard\_name= "Single nucleotide polymorphism"

FT replace(605,r)

FT /\*tag= f

FT /label= RAI-3-s6

FT /notes "exon 2, missense (Ser/Gly)"

FT /standard\_name= "Single nucleotide polymorphism"

FT replace(797,r)

FT /\*tag= g

FT /label= RAI-3-s5

FT /notes "exon 2, missense (Thr/Ala)"

FT /standard\_name= "Single nucleotide polymorphism"

FT replace(1111,r)

FT /\*tag= h

FT /label= RAI-3-s8

FT /notes "silent (Pro/Pro)"

FT /standard\_name= "Single nucleotide polymorphism"

FT replace(1173,r)

FT /\*tag= i

FT /label= RAI-3-s9

FT /note= "missense (Gln/Arg)"  
FT /standard\_name= "Single nucleotide polymorphism"  
XX  
XX WO2004001060-A2.

XX 31-DEC-2003.

XX 20-JUN-2003; 2003WO-US019255.

XX 20-JUN-2002; 2002US-0390850P.

XX 29-AUG-2002; 2002US-0407006P.

XX (BRIM ) BRISTOL-MYERS SQUIBB CO.

XX Whitney GS, Opitck G, Garulacan L, Ramanathan CS, Mckinnon M;  
PI Bennett KL, Barber LE, Cacace A, Tsuchihashi Z;  
XX WPI; 2004-090973/09.

XX P-PSDB; ADI28526.

XX New nucleic acid molecule encoding a human G-protein coupled receptor,

PT RAI3, useful for preventing, treating or ameliorating chronic obstructive  
PT pulmonary disease (COPD), COPD-like disorder, or the underlying symptoms  
PT of COPD.

XX Claim 21; SEQ ID NO 18; 301pp; English.

XX The present sequence is that of cDNA encoding a human G-protein coupled  
PS receptor (Class C, Group 5), denoted retinoic acid induced 3 (RAI-3).  
XX Proteomics methods were used to isolate cigarette smoke-inducible  
CC tyrosine phosphorylated proteins from airway epithelial cells. RAI-3 was  
CC identified as being tyrosine phosphorylated and/or as being  
CC associated/complexed with tyrosine phosphorylated proteins only in those  
CC cells that had been exposed to cigarette smoke. Since RAI-3 is primarily  
CC expressed in lung tissue, and since cigarette smoke is a major causative  
CC factor of chronic obstructive pulmonary disease (COPD), RAI-3 provides a  
CC novel cellular target for identifying modulators, e.g. agonists or  
CC antagonists, useful for the treatment and/or prevention of COPD and  
CC related disorders such as emphysema and chronic bronchitis. RAI-3  
CC modulators, e.g. agonists and antagonists, especially antisense  
CC compounds, can be used to treat COPD and other disorders and diseases  
CC associated with regulation of NF-kB and/or its associated or interacting  
CC signaling molecules. Single nucleotide polymorphisms (SNPs) detected in  
CC the RAI-3 gene are useful for determining COPD association in  
CC individuals. RAI-3 nucleic acid molecules and polypeptides are useful for  
CC preventing, treating or ameliorating disorders related to aberrant GPCR  
CC signaling or cell cycle regulation, pulmonary disorders, inflammatory  
CC lung disorders, COPD, the underlying symptoms of COPD, COPD-related  
CC disorders or conditions, autoimmune disorders, disorders related to  
CC hyperimmune activity, inflammatory conditions, disorders related to  
CC aberrant acute phase responses, hypercongenital conditions, birth  
CC defects, necrotic lesions, wounds, organ transplant rejection, renal  
CC diseases, ischaemia-reperfusion injury, heart disorders, disorders  
CC related to aberrant signal transduction, proliferation disorders,  
CC cancers, HIV infection, asthma, cystic fibrosis, pulmonary fibrosis,  
CC ulcerative colitis, cerebral infarct, myocardial infarct, diabetic  
CC nephropathy, allergic rhinitis, Crohn's disease, atherosclerosis,  
CC rheumatoid arthritis, inflammatory/autoimmune disorders, glioblastoma,  
CC pulmonary small cell undifferentiated carcinoma, carcinoma of the breast,  
CC colon, lung, ovary, pancreas, prostate, non-Hodgkin's lymphoma, disorders  
CC associated with aberrant cell adhesion, I-CAM function and/or regulation,  
CC E-selectin function and/or regulation, or aberrant NF-kB function and/or  
CC regulation (all claimed).

SQ Sequence 2456 BP; 516 A; 719 C; 570 G; 643 T; 0 U; 8 Other;

Query Match 90.4%; Score 30; DB 12; Length 2456;

Best Local Similarity 64.3%; Pred. No. 0.0039;

Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

OY 1 GCNCAYCGNTGGCCNTCCNTAYARGAYTAYGAGTNAAR 42

DB 1271 GCCCAGCGTTGGCGGCCCTTACAAAGACTATGAAGTAAAG 1312

```
RESULT 30
ADI28459
ID ADI28459 standard; cDNA; 2456 BP.
XX AC ADI28459;
XX DT 22-APR-2004 (first entry)
XX DE Human GPCR retinoic acid induced 3 (RAI-3) cDNA.
XX KW Retinoic acid induced 3; RAI-3; human; G-protein coupled receptor; GPCR;
XX KW antiinflammatory; immunosuppressive; cytostatic; cardiant; antiallergic;
XX KW broncholytic; gene therapy; gene; single nucleotide polymorphism; SNP;
XX KW chromosome 12p13-p12.3; 88.
XX OS Homo sapiens.
XX FH Key
XX FT Location/Qualifiers
XX FT replace(112,a)
XX FT /tag= a
XX FT /label= RAI-3-s1
XX FT /note= "located in 5' untranslated region"
XX FT /standard_name= "Single nucleotide polymorphism"
XX FT CDS
XX FT 254..1327
XX FT /tag= b
XX FT /product= "Human RAI3"
XX FT replace(364,t)
XX FT /tag= c
XX FT /label= RAI-3-s2
XX FT /note= "exon 1, silent (Ala/Ala)"
XX FT /standard_name= "Single nucleotide polymorphism"
XX FT replace(51,t)
XX FT /tag= d
XX FT /label= RAI-3-s3
XX FT /note= "exon 2, silent (Ile/Ile)"
XX FT /standard_name= "Single nucleotide polymorphism"
XX FT replace(523,t)
XX FT /tag= e
XX FT /label= RAI-3-s4
XX FT /note= "exon 2, silent (Asp/Asp)"
XX FT /standard_name= "Single nucleotide polymorphism"
XX FT replace(605,g)
XX FT /tag= f
XX FT /label= RAI-3-s6
XX FT /note= "exon 2, missense (Ser/Gly)"
XX FT /standard_name= "Single nucleotide polymorphism"
XX FT replace(797,g)
XX FT /tag= g
XX FT /label= RAI-3-s5
XX FT /note= "exon 2, missense (Thr/Ala)"
XX FT /standard_name= "Single nucleotide polymorphism"
XX FT replace(1111,g)
XX FT /tag= h
XX FT /label= RAI-3-s8
XX FT /note= "silent (Pro/Pro)"
XX FT /standard_name= "Single nucleotide polymorphism"
XX FT replace(1173,g)
XX FT /tag= i
XX FT /label= RAI-3-s9
XX FT /note= "missense (Gln/Arg)"
XX FT /standard_name= "Single nucleotide polymorphism"
XX WO2004001060-A2.
XX 31-DEC-2003.
XX 20-JUN-2003; 2003WO-US019255.
XX 20-JUN-2002; 2002US-0390850P.
XX 29-AUG-2002; 2002US-0407006P.
```

```
PA (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX Whitney GS, Opitck G, Garulacan L, Ramanathan CS, McKinnon M;
PI Bennett KL, Barber LE, Cacace A, Tsuchihashi Z;
XX WPI; 2004-090973/09.
DR P-PSDB; ADI28460.
DR GENBANK; NM_003979.
XX New nucleic acid molecule encoding a human G-protein coupled receptor,
PT RAI3, useful for preventing, treating or ameliorating chronic obstructive
PT pulmonary disease (COPD), COPD-like disorder, or the underlying symptoms
XX of COPD.
PS Claim 1; SEQ ID NO 2; 301pp; English.
XX The present sequence is that of cDNA encoding a human G-protein coupled
CC receptor (Class C, Group 5), denoted retinoic acid induced 3 (RAI-3).
CC Proteomics methods were used to isolate cigarette smoke-inducible
CC tyrosine phosphorylated proteins from airway epithelial cells. RAI-3 was
CC identified as being tyrosine phosphorylated and/or as being
CC associated/complexed with tyrosine phosphorylated proteins only in those
CC cells that had been exposed to cigarette smoke. Since RAI-3 is primarily
CC expressed in lung tissue, and since cigarette smoke is a major causative
CC factor of chronic obstructive pulmonary disease (COPD), RAI-3 provides a
CC novel cellular target for identifying modulators, e.g. agonists or
CC antagonists, useful for the treatment and/or prevention of COPD and
CC related disorders such as emphysema and chronic bronchitis. RAI-3
CC modulators, e.g. agonists and antagonists, especially antisense
CC compounds, can be used to treat COPD and other disorders and diseases
CC associated with regulation of NF-kB and/or its associated or interacting
CC signaling molecules. Single nucleotide polymorphisms (SNPs) detected in
CC the RAI-3 gene are useful for determining COPD association in
CC individuals. RAI-3 nucleic acid molecules and polypeptides are useful for
CC preventing, treating or ameliorating disorders related to aberrant GPCR
CC signaling or cell cycle regulation, pulmonary disorders, inflammatory
CC lung disorders, COPD, the underlying symptoms of COPD, COPD-related
CC disorders or conditions, autoimmune disorders, disorders related to
CC hyperimmune activity, inflammatory conditions, disorders related to
CC aberrant acute phase responses, hypercongenital conditions, birth
CC defects, necrotic lesions, wounds, organ transplant rejection, renal
CC diseases, ischaemia-reperfusion injury, heart disorders, disorders
CC related to aberrant signal transduction, proliferation disorders,
CC cancers, HIV infection, asthma, cystic fibrosis, pulmonary fibrosis,
CC ulcerative colitis, cerebral infarct, myocardial infarct, diabetic
CC nephropathy, allergic rhinitis, Crohn's disease, atherosclerosis,
CC rheumatoid arthritis, inflammatory/autoimmune disorders, glioblastoma,
CC pulmonary small cell undifferentiated carcinoma, carcinoma of the breast,
CC colon, lung, ovary, pancreas, prostate, non-Hodgkin's lymphoma, disorders
CC associated with aberrant cell adhesion, I-CAM function and/or regulation,
CC E-selectin function and/or regulation, or aberrant NF-kB function and/or
CC regulation (all claimed).
XX Sequence 2456 BP; 519 A; 722 C; 571 G; 644 T; 0 U; 0 Other;
SQ
Query Match 90.4%; Score 30; DB 12; Length 2456;
Best Local Similarity 64.3%; Pred. No. 0.0039;
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;
Oy 1 GCNCAYGCGTGGCCGAGCCCTTACAGACTATGAGTAAG 1312
Db 1271 GCCCAGCTGGCCGAGCCCTTACAGACTATGAGTAAG 1312
RESULT 31
ADI28459
ID ADR48221 standard; cDNA; 2456 BP.
XX AC ADR48221;
XX DT 18-NOV-2004 (first entry)
XX DE Human retinoic acid induced 3 (RAI3) encoding cDNA SEQ:9.
```





ID ADR43992 standard; DNA; 2456 BP.  
AC ADR43992;  
XX  
DT 18-NOV-2004 (first entry)  
XX  
DE Human breast tumour associated gene clone-02 SEQ ID NO:2.  
XX  
KW cancer; tumour; breast cancer; lung cancer; colon cancer; kidney cancer;  
KW antibody; cytostatic; gene therapy; human; breast tumour tissue;  
KW breast tissue; gene; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO2004074506-A2.  
XX  
PD 02-SEP-2004.  
XX  
PF 30-JAN-2004; 2004WO-US003030.  
XX  
PR 13-FEB-2003; 2003US-0447900P.  
XX  
PA (MERG-) MERGEN LTD.  
XX  
PI Hu Q, Peng A, Liu B, Love JR, Hao X, Ren M, Sheng Z;  
XX  
DR WPI; 2004-635589/61.  
XX  
XX Diagnosing and treating a cancer, e.g. breast, lung, colon, or kidney  
PT cancer, comprises detecting and blocking the over expression of a gene of  
PT a protein found in breast, lung, colon, or kidney tissue.  
XX  
PS Claim 2; SEQ ID NO 2; 220pp; English.  
XX  
CC The present invention describes a method for diagnosing and treating a  
CC cancer, e.g. breast, lung, colon, or kidney cancer. The method comprises  
CC detecting and blocking the over expression of a gene of a protein found  
CC in breast, lung, colon, or kidney tissue. Also described is an antibody  
CC or binding portion of an antibody that specifically binds a protein found  
CC in breast tissue, lung tissue, colon tissue, or kidney tissue. The  
CC antibody has cytostatic activity, and can be used in gene therapy. The  
CC methods, antibodies, polynucleotides and polypeptides from the present  
CC invention are useful for detecting, diagnosing, preventing and treating  
CC cancer, e.g. breast, lung, colon, or kidney cancer. The present sequence  
CC represents a nucleotide sequence given in the present invention, which is  
CC over expressed in human breast tumour tissue. N.B. All 385 sequences  
CC referenced in this patent are detailed in the US provisional application  
CC SN 60/447,900, filed 02/13/2003. In this application only one sequence,  
CC representing the longest sequence of each of the 65 clones is listed in  
CC Appendix A (and given as SEQ ID NO:1 to 65 in the Sequence Listing).  
XX  
SQ Sequence 2456 BP; 517 A; 710 C; 567 G; 643 T; 0 U; 19 Other;  
Query Match 90.4%; Score 30; DB 13; Length 2456;  
Best Local Similarity 64.3%; Pred. No. 0.0039;  
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;  
Qy 1 GCNCAYGNTGCCNTCCNTAYAAAGCTATGAAGTAAAG 42  
Db 1271 GCCCAGCTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1312  
RESULT 34  
ADU06126  
ID ADU06126 standard; DNA; 2456 BP.  
XX  
AC ADU06126;  
XX  
DT 27-JAN-2005 (first entry)  
XX  
DE Novel bronchial cancer-associated human gene SeqID350.  
XX  
KW bronchial cancer; cytostatic; tumour-associated protein;  
cancer detection; metastasis; tumour; gene; ds; human.  
XX Homo sapiens.  
XX DE10316701-A1.  
XX 04-NOV-2004.  
XX 09-APR-2003; 2003DE-01016701.  
XX 09-APR-2003; 2003DE-01016701.  
XX (HINZ/) HINZMANN B.  
XX (HERM/) HERMANN K.  
XX (CAST/) HEIDEN CASTANOS-VELEZ E.  
XX Mennerich D, Bruemendorf T, Heiden E, Hermann K, Kinnemann H;  
PI Li X, Roepcke S, Staub E, Hinzmann B, Rosenthal A, Pillarsky C;  
XX WPI; 2004-786403/78.  
DR P-PSDB; ADU06613.  
XX  
XX New nucleic acid, and derived proteins, useful for diagnosis of bronchial  
PT cancer and in screening for therapeutic and diagnostic agents.  
XX  
PS Claim 1; SEQ ID NO 350; 1381pp; German.  
XX  
CC This invention relates to a novel isolated nucleic acid associated with  
CC bronchial cancer comprising 489 defined sequences given in the  
CC specification. The invention may be useful for the production of  
CC compounds with a cytostatic activity through the inhibition of expression  
CC or activity of tumour-associated proteins. The novel DNA sequences and  
CC the proteins/peptides encoded by them are used for detecting bronchial  
CC cancer or determining the risk of developing it and to screen for  
CC specific binding partners of the DNA or protein sequences, where the  
CC binding partners are potentially useful as agents for treating or  
CC diagnosing bronchial cancer. The DNA or protein sequences can also be  
CC used for prognosis, detection of metastases and for secondary treatment  
CC (of tumours that have been stabilised or are no longer detectable).  
CC Detecting abnormal expression of the DNA sequences provides early  
CC diagnosis of bronchial cancers. The present sequence is that of a novel  
CC bronchial cancer-associated human gene sequence of the invention.  
XX  
SQ Sequence 2456 BP; 519 A; 722 C; 571 G; 644 T; 0 U; 0 Other;  
Query Match 90.4%; Score 30; DB 13; Length 2456;  
Best Local Similarity 64.3%; Pred. No. 0.0039;  
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;  
Qy 1 GCNCAYGNTGCCNTCCNTAYAAAGCTATGAAGTAAAG 42  
Db 1271 GCCCAGCTTGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1312  
RESULT 35  
AEA00088  
ID AEA00088 standard; cDNA; 2456 BP.  
XX  
AC AEA00088;  
XX  
DT 28-JUL-2005 (first entry)  
XX  
DE Human TAT115 cDNA SEQ ID NO:40.  
XX  
XX ss; gene; tumor-associated antigen; cytostatic; breast tumor;  
KW endocrine disease; gynecology and obstetrics; neoplasm; colon tumor;  
KW gastrointestinal disease; rectal tumor; endometroid carcinoma;  
KW genitourinary disease; renal tumor; lung tumor; respiratory disease;  
KW ovary tumor; skin tumor; liver tumor.  
XX  
OS Homo sapiens.  
XX  
XX Key Location/Qualifiers  
FH









```
XX AC ADL13501;
XX DT 06-MAY-2004 (first entry)
XX DE Osteoarthritis-associated polymorphic nucleotide #33.
XX KW ds; gene; osteopathic; antiinflammatory; antiarthritic; gene therapy;
XX KW joint space narrowing; osteophyte development; joint pain;
XX KW osteoarthritis; SNP; single nucleotide polymorphism.
XX OS Homo sapiens.
XX PN WO2003054166-A2.
XX PD 03-JUL-2003.
XX PF 19-DEC-2002; 2002WO-US041225.
XX XX 20-DEC-2001; 2001US-0342603P.
XX PA (INCY-) INCYTE GENOMICS INC.
XX PI Jones KA, Schafer A;
XX XX WPI; 2003-559141/52.
XX DR
XX PT Determining susceptibility of an individual to joint space narrowing,
XX PT osteophyte development and/or joint pain comprises identifying whether
XX PT the individual has at least one polymorphism in a polynucleotide encoding
XX PT a protein.
XX PS Disclosure; SEQ ID NO 33; 287bp; English.
XX CC
XX CC The invention relates to a method of determining susceptibility of an
XX CC individual to joint space narrowing and/or osteophyte development and/or
XX CC joint pain comprising identifying whether the individual has at least one
XX CC polymorphism in a polynucleotide encoding at least one of the protein
XX CC listed in the specification. The methods, composition and agent are
XX CC useful for modulating the susceptibility of an individual to joint space
XX CC narrowing and/or osteophyte development and/or joint pain that is
XX CC associated with a disease, preferably osteoarthritis. The cell line and
XX CC the non-human animal are useful for screening for an agent for diagnosing
XX CC an individual having susceptibility to joint space narrowing and/or
XX CC osteophyte development and/or joint pain. This sequence corresponds to
XX CC the polynucleotide encoding a protein listed in the specification. (Note:
XX CC The sequence data for this patent did not form part of the printed
XX CC specification but was obtained in electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences).
XX SQ
XX Query Match 64.5%; Score 21.4; DB 10; Length 167932;
XX Best Local Similarity 57.9%; Pred. No. 1.2e+02;
XX Matches 22; Conservative 5; Mismatches 11; Indels 0; Gaps 0;
XX
XX QY 1 GCNAYGCGTCCNCNTAYAAAGAYTAYGAGT 38
XX |||||
XX Db 130330 GCACACTCATTCTTCCTCTTAAAAAATCTATGAGT 130293
XX
XX RESULT 44
XX ADR84444
XX ID ADR84444 standard; DNA; 7880 BP.
XX AC ADR84444;
XX AC
XX DT 04-NOV-2004 (first entry)
XX DE Aspergillus fumigatus essential gene genomic sequence #255.
XX KW Fungicide; Aspergillus fumigatus infection; Farmer's lung disease;
XX KW drug screening; ds.
XX
```

```
XX OS Aspergillus fumigatus.
XX PN WO2004067709-A2.
XX XX
XX PD 12-AUG-2004.
XX XX
XX PF 16-JAN-2004; 2004WO-US001099.
XX XX 17-JAN-2003; 2003US-0441281P.
XX PR 13-JUN-2003; 2003US-0478196P.
XX XX
XX PA (ELIT-) ELITRA PHARM INC.
XX PA (ELIT-) ELITRA CANADA LTD.
XX PI Jiang B, Hu W, Lemieux S, Roemer T;
XX XX
XX DR WPI; 2004-594200/57.
XX DR P-PSDB; ADR86205.
XX XX
XX PT New purified or isolated Aspergillus fumigatus nucleic acid molecule
XX PT encoding a gene product, useful for diagnosing and/or treating invasive
XX PT fungal infections, such as Farmer's lung disease.
XX XX
XX PS Claim 3; SEQ ID NO 255; 164pp; English.
XX XX
XX CC The present invention relates to Aspergillus fumigatus genes that are
XX CC essential and are potential targets for drug screening. The methods and
XX CC compositions of the present invention are useful for diagnosing and/or
XX CC treating invasive Aspergillus fumigatus infection, including the allergic
XX CC forms of the disease, such as Farmer's lung disease. They can also be
XX CC used in various drug discovery purposes, such as expression of the
XX CC recombinant protein, hybridization assay and construction of nucleic acid
XX CC arrays. The present sequence represents an Aspergillus fumigatus
XX CC essential gene full length genomic sequence, used during diagnosis and
XX CC drug development in the invention. These genes share a high degree of
XX CC sequence conservation with known essential genes of candida albicans. The
XX CC sequence data for this patent is not represented in the printed
XX CC specification, but was obtained in electronic format from WIPO.
XX XX
XX SQ Sequence 7880 BP; 1905 A; 1902 C; 1920 G; 2153 T; 0 U; 0 Other;
XX Query Match 63.9%; Score 21.2; DB 13; Length 7880;
XX Best Local Similarity 57.1%; Pred. No. 62;
XX Matches 20; Conservative 6; Mismatches 9; Indels 0; Gaps 0;
XX
XX QY 6 YCGNTGGCCNTCCNTAYAAAGAYTAYGAGTNA 40
XX :|||:|||||:|||||:|||||:|||||:|||||:
XX Db 1152 CGCGTGGCCCTCTCATATCAGGATCATGCTTA 1186
XX
XX RESULT 45
XX ADO56529
XX ID ADO56529 standard; DNA; 421 BP.
XX XX
XX AC ADO56529;
XX XX
XX DT 21-OCT-2004 (first entry)
XX XX
XX DE Novel canine microarray-related DNA sequence SeqID7831.
XX XX
XX KW canine microarray; drug screening; toxicity assay;
XX KW environmental pollutant; cellular response; gene expression profile;
XX KW toxic response; liver necrosis; fatty liver disease;
XX KW protein adduct formation; hepatitis; dog; ds.
XX XX
XX OS Canis familiaris.
XX XX
XX PN WO2004063324-A2.
XX XX
XX PD 29-JUL-2004.
XX XX
XX PF 05-MAY-2003; 2003WO-US013853.
```

XX	03-MAY-2002; 2002US-0377240P.
PR	(GENE-) GENE LOGIC INC.
XX	(PFIZ ) PFIZER PROD INC.
PA	Diggins JC, Porter M, Wei T;
PI	WPI; 2004-561890/54.
XX	New isolated nucleic acid molecule, useful for drug screening and
XX	toxicity assays or for assessing the impact, including toxicity, of a
PT	compound, pharmaceutical agent or environmental pollutant on a cell or
PT	living organism.
XX	Claim 1; SEQ ID NO 7831; 4lpp; English.
PS	This invention is related to a novel isolated canine nucleic acid
XX	sequences and the construction of canine microarrays containing a
CC	significant portion of the canine genome. The isolated canine nucleic
CC	acid sequences of the invention may be useful for drug screening and
CC	toxicity assays. The invention is therefore useful for assessing the
CC	impact, including toxicity, of a compound, pharmaceutical agent or
CC	environmental pollutant on a cell or living organism. The methods are
CC	useful for detecting genes that are up- or down-regulated in canines in a
CC	disease state. The sequences are useful as diagnostic agents or markers
CC	to detect a cellular response in a sample individually or as part of a
CC	gene expression profile. It is also useful as a target for agents that
CC	modulate gene expression or activity. The database is useful for
CC	producing electronic Northern blots that allow the user to determine the cell
CC	type or tissue in which a given gene is expressed and to allow
CC	determination of the abundance or expression level of a given gene in a
CC	particular tissue or cell. The methods are useful for determining the
CC	similarity of a toxic response to one or more individual compounds. The
CC	methods are useful for predicting at least one toxic response or the
CC	likelihood that a compound or test agent will induce various specific
CC	pathologies such as those of the liver (liver necrosis), fatty liver
CC	disease, protein adduct formation or hepatitis), those of the kidney,
CC	heart, brain or testes, or other pathologies associated with at least one
CC	of the toxins. The methods are also useful for predicting or elucidating
CC	the potential cellular pathways influenced, induced or modulated by the
CC	compound or test agent due to the similarity of the expression profile
CC	compared to the profile induced by a known toxin. The present sequence is
CC	that of a canine DNA sequence which was claimed for use during the
CC	production of a canine microarray of the invention.
XX	Sequence 421 BP; 59 A; 147 C; 92 G; 87 T; 0 U; 36 Other;
SQ	
Query Match	62.7%; Score 20.8; DB 13; Length 421;
Best Local Similarity	55.0%; Pred. No. 41;
Matches	22; Conservative 5; Mismatches 13; Indels 0; Gaps 0;
QY	1 GCNCAYGCVTGGCCNTCCNTAATGAAGTAATGARGTNA 40
DB	301 GCCCATCGAGGCCCTCACACCAAGGCCTACGCTCTTA 340
RESULT 46	
ABQ81845/c	
ID	ABQ81845 standard; DNA; 349980 BP.
XX	ABQ81845;
AC	
XX	19-NOV-2002 (first entry)
DT	
DE	Bifidobacterium longum NCC2705 related nucleotide sequence SEQ ID:1101.
XX	
KW	Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;
KW	antidiarrheic; antibacterial; inhibitor of Salmonella; detection;
KW	identification; lactic acid bacterium; diarrhoea; pathogenic bacteria;
XX	rotavirus; food composition; pharmaceutical composition; gene; ds.
XX	
OS	Bifidobacterium longum.
XX	
OS	Synthetic.
XX	EP1227152-A1.
PN	31-JUL-2002..
XX	30-JAN-2001; 2001EP-00102050.
PF	30-JAN-2001; 2001EP-00102050.
PR	(NEST ) SOC PROD NESTLE SA.
PA	WPI; 2002-668997/72.
XX	Novel polynucleotide comprising Bifidobacterium genome sequence useful as
PT	a probe or primer for detecting and/or identifying Bifidobacterium longum
PT	in a biological sample.
XX	Disclosure; SEQ ID NO 1101; 80pp; English.
PS	The present invention describes a polynucleotide (I) comprising a
CC	sequence of a Bifidobacterium genome selected from the nucleotide
CC	sequences given in ABQ81842 and ABQ81843, or a sequence exhibiting at
CC	least 90% identity or which hybridises with the sequences given in
CC	ABQ81842 and ABQ81843. Also described is a polynucleotide (II) encoding a
CC	fusion protein, comprising a sequence selected from 1097 sequences given
CC	in ABP65258 to ABP66354 ligated in frame to a polynucleotide encoding a
CC	heterologous polypeptide. (I) has antidiarrheic and antibacterial
CC	activities, and can be used as an inhibitor of Salmonella. (I) (which is
CC	a probe) is useful for the detection and/or identification of
CC	Bifidobacterium longum in a biological sample. A carrier containing the
CC	lactic acid bacterium Bifidobacterium longum NCC2705 (CNCM I-2618) can be
CC	used for preventing and/or treating diarrhoea brought about by pathogenic
CC	bacteria and/or rotavirus. The carrier is a food composition selected
CC	from milk, yogurt, curd, cheese, fermented milks, milk based fermented
CC	products, ice-creams, fermented cereal based products, milk based
CC	powders, infant formula, pet food or a pharmaceutical composition
CC	selected from tablets, liquid bacterial suspensions, dried oral
CC	supplement, wet oral supplement, dry tube feeding or wet tube feeding.
CC	(I) is useful in DNA arrays or chips to carry out analysis of the
CC	expression of the Bifidobacterium gene. ABQ81844 to ABQ81850 represent
CC	Bifidobacterium related nucleotide sequences given in the Sequence
CC	Listing from the present invention but not mentioned further within the
CC	specification. N.B. The sequence data for this patent is not represented
CC	in the printed specification but is based on sequence information
CC	supplied by the European Patent Office
XX	Sequence 349980 BP; 70780 A; 106600 C; 104724 G; 67876 T; 0 U; 0 Other;
SQ	
Query Match	62.7%; Score 20.8; DB 6; Length 349980;
Best Local Similarity	55.0%; Pred. No. 2.7e+02;
Matches	22; Conservative 5; Mismatches 13; Indels 0; Gaps 0;
QY	1 GCNCAYGCVTGGCCNTCCNTAATGAAGTAATGARGTNA 40
DB	129737 GCCCACCACACTGTCCCTCCAGTAGAAGGACTGCGAGCGCA 129698
RESULT 47	
AAT51538	
ID	AAT51538 standard; DNA; 2004 BP.
XX	AAT51538;
AC	
XX	16-OCT-2003 (revised)
DT	23-APR-1997 (first entry)
DE	Mycobacterium gallisepticum incomplete pMGAL.5/1.6 gene.
XX	
KW	Adhesin; pMGAL.5; mycoplasma; diagnosis; vaccine; vector;
KW	respiratory disease; poultry; haemagglutinin; promoter; probe; ds.
XX	
OS	Mycoplasma gallisepticum; strain S6.

```
XX CA2135330-A.
PN
XX
XX 11-MAY-1995.
XX
XX 08-NOV-1994; 94CA-02135330.
XX
XX 10-NOV-1993; 93AU-00050593.
PR 20-APR-1994; 94US-00230312.
XX
XX (BROW/) BROWNING G F.
XX
XX Browning GF, Markham PF, Whithear KG, Walker ID, Glew MD;
XX WPI; 1995-241027/32.
DR
XX
XX New promoter region from a Mycoplasma gallisepticum adhesin gene - useful
PT when coupled to foreign antigen gene, for prodn. of multivalent live
PT vaccines, also new probes for detecting Mycoplasma and manipulating its
PT genome.
XX
XX Claim 2; Page 46-48; 81pp; English.
XX
XX A 10 kb DNA fragment was isolated from a Mycoplasma gallisepticum genomic
CC library using probes (AAT51533-34) based on tryptic peptides of a M.
CC gallisepticum PMGA adhesin. The fragment was divided into 5 putative
CC genes, PMGA1.2, PMGA1.3, PMGA1.4 and PMGA1.5/1.6 (AAT51535-38) on the
CC basis of deduced amino acids sequences. The 5 genes are closely related.
CC The regions between the genes (see also AAT51539-42) are highly
CC conserved. A consensus conserved sequence (AAT51532) can be used as a
CC probe to detect Mycoplasma infection, esp. in poultry. The promoter
CC sequences found in the conserved sequences may be linked to exogenous
CC antigen genes for prodn. of multivalent live vaccines. (Updated on 16-OCT
CC -2003 to standardise OS field)
XX
XX Sequence 2004 BP; 756 A; 317 C; 335 G; 596 T; 0 U; 0 Other;
SQ
Query Match 62.0%; Score 20.6; DB 2; Length 2004;
Best Local Similarity 52.6%; Pred. No. 79;
Matches 20; Conservative 7; Mismatches 11; Indels 0; Gaps 0;
QY 5 AYCGNTGGCCNTCCNTAYARGAYTAYGARGTNAAR 42
Db 818 ATGCTGGACAATCACCTAATAAGATAATGTAGTAAAG 855
RESULT 48
AAQ68672
ID AAQ68672 standard; DNA; 2577 BP.
XX
XX AAQ68672;
AC
XX
XX 28-FEB-1995 (first entry)
DT
XX
XX Partial PMGA1.5 and partial PMGA1.6 Mycoplasma gallisepticum DNA.
DE
XX
XX PMGA; adhesin gene complex; hemagglutinin; conserved sequences; primers;
KW probes; amplification; polymerase chain reaction; specific; detection;
KW PCR; 1.2; 1.3; 1.4; 1.5; 1.6; ss.
XX
XX Mycoplasma gallisepticum.
OS
XX
XX Key Location/Qualifiers
FH misc_feature 1. .1333
FT /tag= C
FT /note= "5'end of PMGA1.5"
FT misc_feature 61
FT /tag= a
FT /note= "unidentified residue"
FT misc_feature 1334. .1857
FT /tag= b
FT /note= "unidentified residues"
FT misc_feature 1857. .2557
```

```
FT /tag= d
FT /note= "3'end of PMGA1.6"
XX
XX AU9350593-A.
PN
XX 26-MAY-1994.
XX
XX 10-NOV-1993; 93AU-00050593.
XX 10-NOV-1992; 92AU-00005744.
PR
XX (UYME ) UNIV MELBOURNE.
PA
XX Browning GF, Markham PF, Whithear KG, Walker ID, Glew MD;
PI WPI; 1994-209061/26.
XX
XX Recombinant DNA constructs for Mycoplasma gallisepticum - for diagnosis,
PT treatment and prophylaxis of poultry respiratory disorders.
PT
XX Example 1; Fig 4; 51pp; English.
PS
XX AAQ68672 shows the 5'end of PMGA1.5 and the 3'end of PMGA1.6 DNA
CC sequences derived from a 10 kb insert from a Mycoplasma gallisepticum
CC genomic DNA library, detected by probes based on the T3 and C7 peptides.
CC Mycoplasma gallisepticum infection in poultry, humans and other animals
CC is of economic importance to many industries and it is desirable to
CC produce effective vaccines and probes for its detection. The sequences
CC and probes and vaccine vectors of the invention can be used for the
CC diagnosis and treatment of Mycoplasma gallisepticum infection, and for
CC prophylaxis
XX
XX Sequence 2577 BP; 772 A; 327 C; 340 G; 613 T; 0 U; 525 Other;
SQ
Query Match 62.0%; Score 20.6; DB 2; Length 2577;
Best Local Similarity 52.6%; Pred. No. 85;
Matches 20; Conservative 7; Mismatches 11; Indels 0; Gaps 0;
QY 5 AYCGNTGGCCNTCCNTAYARGAYTAYGARGTNAAR 42
Db 817 ATGCTGGACAATCACCTAATAAGATAATGTAGTAAAG 854
RESULT 49
AED71703
ID AED71703 standard; DNA; 1053 BP.
XX
XX AED71703;
AC
XX
XX 12-JAN-2006 (first entry)
DT
XX
XX Corynebacterium glutamicum MCP gene SEQ ID NO:1879.
DE
XX
XX marker and fine chemical production protein; screening; gene; ds.
KW
XX
XX Corynebacterium glutamicum.
OS
XX
XX US6962989-B1.
PN
XX 08-NOV-2005.
PD
XX
XX 27-JUN-2000; 2000US-00605703.
PF
XX
XX 08-JUL-1999; 99US-0142764P.
PR 03-SEP-1999; 99US-0152318P.
XX
XX (BADI ) BASF AG.
PA
XX
XX Pompejus M, Kroegeer B, Schroeder H, Zelder O, Haberhauer G;
PI WPI; 2005-743593/76.
XX
XX P-PSDB; AED71704.
DR
XX
```









WP AEB39175\_22 2200001 2310000  
WP AEB39175\_23 2300001 2410000  
WP AEB39175\_24 2400001 2510000  
WP AEB39175\_25 2500001 2610000  
WP AEB39175\_26 2600001 2710000  
WP AEB39175\_27 2700001 2810000  
WP AEB39175\_28 2800001 2910000  
WP AEB39175\_29 2900001 3010000  
WP AEB39175\_30 3000001 3110000  
WP AEB39175\_31 3100001 3210000  
WP AEB39175\_32 3200001 3310000  
WP AEB39175\_33 3300001 3410000  
WP AEB39175\_34 3400001 3503610

Query Match 60.8%; Score 20.2; DB 14; Length 110000;  
Best Local Similarity 53.7%; Pred. No. 3.7e+02;  
Matches 22; Conservative 5; Mismatches 14; Indels 0; Gaps 0;

QY 1 GCNCAYGCGTGGCCNTCCNTAYAGAYTAYGAGTNA 41  
DB 66257 GCCATAAGTTCGCTGCTCTGCTATAAATAGGACGAAGCAAA 66297

## RESULT 57

AEB42401\_22  
Continuation (23 of 34) of AEB42401 from base 2200001 (L. pneumophila DNA SEQ ID NO 6733)  
WP Sequence split into 34 fragments LOCUS AEB42401 Accession Aeb42401

WP Fragment Name Begin End  
WP AEB42401\_00 1 110000  
WP AEB42401\_01 100001 210000  
WP AEB42401\_02 200001 310000  
WP AEB42401\_03 300001 410000  
WP AEB42401\_04 400001 510000  
WP AEB42401\_05 500001 610000  
WP AEB42401\_06 600001 710000  
WP AEB42401\_07 700001 810000  
WP AEB42401\_08 800001 910000  
WP AEB42401\_09 900001 1010000  
WP AEB42401\_10 1000001 1110000  
WP AEB42401\_11 1100001 1210000  
WP AEB42401\_12 1200001 1310000  
WP AEB42401\_13 1300001 1410000  
WP AEB42401\_14 1400001 1510000  
WP AEB42401\_15 1500001 1610000  
WP AEB42401\_16 1600001 1710000  
WP AEB42401\_17 1700001 1810000  
WP AEB42401\_18 1800001 1910000  
WP AEB42401\_19 1900001 2010000  
WP AEB42401\_20 2000001 2110000  
WP AEB42401\_21 2100001 2210000  
WP AEB42401\_22 2200001 2310000  
WP AEB42401\_23 2300001 2410000  
WP AEB42401\_24 2400001 2510000  
WP AEB42401\_25 2500001 2610000  
WP AEB42401\_26 2600001 2710000  
WP AEB42401\_27 2700001 2810000  
WP AEB42401\_28 2800001 2910000  
WP AEB42401\_29 2900001 3010000  
WP AEB42401\_30 3000001 3110000  
WP AEB42401\_31 3100001 3210000  
WP AEB42401\_32 3200001 3310000  
WP AEB42401\_33 3300001 3345687

Query Match 60.8%; Score 20.2; DB 14; Length 110000;  
Best Local Similarity 53.7%; Pred. No. 3.7e+02;  
Matches 22; Conservative 5; Mismatches 14; Indels 0; Gaps 0;

QY 1 GCNCAYGCGTGGCCNTCCNTAYAGAYTAYGAGTNA 41  
DB 39753 GCCATAAGTTCGCTGCTCTGCTATAAATAGGACGAAGCAAA 39793

## RESULT 58

## AEB42737\_08

Continuation (9 of 21) of AEB42737 from base 800001 (L. pneumophila DNA SEQ ID NO 7069. )  
WP Sequence split into 21 fragments LOCUS AEB42737 Accession Aeb42737

WP Fragment Name Begin End  
WP AEB42737\_00 1 110000  
WP AEB42737\_01 100001 210000  
WP AEB42737\_02 200001 310000  
WP AEB42737\_03 300001 410000  
WP AEB42737\_04 400001 510000  
WP AEB42737\_05 500001 610000  
WP AEB42737\_06 600001 710000  
WP AEB42737\_07 700001 810000  
WP AEB42737\_08 800001 910000  
WP AEB42737\_09 900001 1010000  
WP AEB42737\_10 1000001 1110000  
WP AEB42737\_11 1100001 1210000  
WP AEB42737\_12 1200001 1310000  
WP AEB42737\_13 1300001 1410000  
WP AEB42737\_14 1400001 1510000  
WP AEB42737\_15 1500001 1610000  
WP AEB42737\_16 1600001 1710000  
WP AEB42737\_17 1700001 1810000  
WP AEB42737\_18 1800001 1910000  
WP AEB42737\_19 1900001 2010000  
WP AEB42737\_20 2000001 2017010

Query Match 60.8%; Score 20.2; DB 14; Length 110000;  
Best Local Similarity 53.7%; Pred. No. 3.7e+02;  
Matches 22; Conservative 5; Mismatches 14; Indels 0; Gaps 0;

QY 1 GCNCAYGCGTGGCCNTCCNTAYAGAYTAYGAGTNA 41  
DB 62418 GCCATAAGTTCGCTGCTCTGCTATAAATAGGACGAAGCAAA 62458

## RESULT 59

AEB35718/C  
ID AEB35718 standard; DNA; 189430 BP.

XX AEB35718;  
XX 08-SEP-2005 (first entry)  
XX L. pneumophila DNA SEQ ID NO 50.  
XX detection; infection; Antibacterial; Vaccine; ds; gene.  
XX Legionella pneumophila.  
XX WO2005049642-A2.  
XX 02-JUN-2005.  
XX 23-SEP-2004; 2004WO-IB003578.  
XX 21-NOV-2003; 2003FR-00013687.  
XX (INSP ) INST PASTEUR.  
XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
XX (UYLY-) UNIV LYON 1 BERNARD CLAUDE.  
XX (CNRS ) CNRS CENT NAT RECH SCI.

PI Buchrieser C, Tichit M, Etienne J, Ma L, Cazalet C, Glaser P;  
PI Rusniok C, Bouchier C, Zidane N, Magnier A, Kunst F, Vandenesch F;  
PI Jarraud S;  
XX WPI; 2005-388305/40.

XX New genome of Legionella pneumophila Paris strain and derived  
PT polypeptides, useful for detection or identification of the strain and  
PT for treatment and prevention of infections.

XX Claim 1; SEQ ID NO 50; 660pp; English.

XX The invention relates to an isolated or purified nucleotide sequences (I)  
CC from Legionella pneumophila Paris strain (II), and their related  
CC sequences or fragments, are useful as primers and probes for detection  
CC and amplification, including differentiation between the Paris and  
CC Philadelphia strains of Legionella pneumophila and to prepare recombinant  
CC (hybrid) polypeptides (III). (II) are also useful for preparation of  
CC specific antibodies (Ab), also used for detection/identification of  
CC Legionella, and some (I), specifically those involved in synthesis of  
CC surface proteins, are targets for identification of inhibitors. (II), or  
CC vectors that contain (I), are useful as vaccines and immunogenic  
CC compositions, for treatment and prevention of infections by L.  
CC pneumophila. The present sequence represents a L. pneumophila DNA.  
XX  
SQ Sequence 189430 BP; 56051 A; 34903 C; 40210 G; 58266 T; 0 U; 0 Other;  
Query Match 60.8%; Score 20.2; DB 14; Length 189430;  
Best Local Similarity 53.7%; Pred. No. 4.3e+02;  
Matches 22; Conservative 5; Mismatches 14; Indels 0; Gaps 0;  
QY 1 GCNCAYCGTGGCCNCCNTAYARGAYTAYGAGTNA 41  
DB 129266 GCCCATAAGTTGCGCTGCTGCCTATAAATAGGACGAGCAA 129226  
RESULT 60  
AEB39171  
ID AEB39171 standard; DNA; 263853 BP.  
XX  
AC AEB39171;  
XX  
DT 08-SEP-2005 (first entry)  
DE L. pneumophila DNA SEQ ID NO 3503.  
XX  
KW detection; infection; Antibacterial; Vaccine; ds; gene.  
XX  
OS Legionella pneumophila.  
XX  
PN WO2005049642-A2.  
XX  
PD 02-JUN-2005.  
XX  
PF 23-SEP-2004; 2004WO-IB003578.  
XX  
PR 21-NOV-2003; 2003FR-00013687.  
XX  
PA (INSP ) INST PASTEUR.  
PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
PA (UPLY-) UNIV LYON 1 BERNARD CLAUDE.  
PA (CNRS ) CNRS CENT NAT RECH SCI.  
XX  
PI Buchrieser C, Tichit M, Etienne J, Ma L, Cazalet C, Glaser P;  
PI Rusniok C, Bouchier C, Zidane N, Magnier A, Kunst F, Vandenesch F;  
PI Jarraud S;  
XX  
WPI; 2005-388305/40.  
XX  
PT New genome of Legionella pneumophila Paris strain and derived  
PT polypeptides, useful for detection or identification of the strain and  
PT for treatment and prevention of infections.  
XX  
PS Example 9; SEQ ID NO 3503; 660pp; English.  
XX  
CC The invention relates to an isolated or purified nucleotide sequences (I)  
CC from Legionella pneumophila Paris strain. (II), and their related  
CC sequences or fragments, are useful as primers and probes for detection  
CC and amplification, including differentiation between the Paris and  
CC Philadelphia strains of Legionella pneumophila and to prepare recombinant  
CC (hybrid) polypeptides (III). (II) are also useful for preparation of  
CC specific antibodies (Ab), also used for detection/identification of  
CC Legionella, and some (I), specifically those involved in synthesis of  
CC surface proteins, are targets for identification of inhibitors. (II), or

CC vectors that contain (I), are useful as vaccines and immunogenic  
CC compositions, for treatment and prevention of infections by L.  
CC pneumophila. The present sequence represents a L. pneumophila DNA.  
XX  
SQ Sequence 263853 BP; 83002 A; 52487 C; 46521 G; 81843 T; 0 U; 0 Other;  
Query Match 60.8%; Score 20.2; DB 14; Length 263853;  
Best Local Similarity 53.7%; Pred. No. 4.7e+02;  
Matches 22; Conservative 5; Mismatches 14; Indels 0; Gaps 0;  
QY 1 GCNCAYCGTGGCCNCCNTAYARGAYTAYGAGTNA 41  
DB 162843 GCCCATAAGTTGCGCTGCTGCCTATAAATAGGACGAGCAA 162883  
RESULT 61  
ABL91353/c  
ID ABL91353 standard; DNA; 669 BP.  
XX  
AC ABL91353;  
XX  
DT 29-AUG-2003 (revised)  
DT 29-JUL-2002 (first entry)  
XX  
DE Chlamydia pneumoniae cp7041 ORF DNA, SEQ ID NO:340.  
XX  
KW Chlamydial infection; antigen; immunogen; vaccine; diagnosis;  
KW human respiratory disease; cardiovascular disease; atherosclerosis;  
KW coronary artery disease; carotid artery stenosis; myocardial infarction;  
KW cerebrovascular disease; aortic aneurysm; claudication; stroke;  
KW strain CWL029; open reading frame; ORF; gene; ds.  
XX  
OS Chlamydia pneumoniae.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..669  
FT /tag= a  
FT /product= "cp7041"  
XX  
PN WO200202606-A2.  
XX  
PD 10-JAN-2002.  
XX  
PF 03-JUL-2001; 2001WO-IB001445.  
XX  
PR 03-JUL-2000; 2000GB-00016363.  
PR 11-JUL-2000; 2000GB-00017047.  
PR 21-JUL-2000; 2000GB-00017983.  
PR 07-AUG-2000; 2000GB-00019368.  
PR 18-AUG-2000; 2000GB-00020440.  
PR 14-SEP-2000; 2000GB-00022583.  
PR 10-NOV-2000; 2000GB-00027549.  
PR 22-DEC-2000; 2000GB-00031706.  
XX  
PA (CHIR-) CHIRON SPA.  
XX  
PI Ratti G, Grandi G;  
XX  
WPI; 2002-154726/20.  
DR N-PSDB; ABB90695.  
XX  
PT Novel Chlamydia pneumoniae protein useful in the manufacture of a  
PT medicament for treatment or prevention of infection due to Chlamydia,  
PT preferably Chlamydia pneumoniae, and for diagnostic purposes.  
XX  
PS Claim 5; Page 176; 364pp; English.  
XX  
CC Sequences ABB90526-ABB90715 represent novel proteins from Chlamydia  
CC pneumoniae (strain CWL029), and ABL91184-ABL91373 represent DNA encoding  
CC them. The proteins are predicted to be immunogenic and may therefore be  
CC useful in vaccine production and for diagnostic purposes. Chlamydia  
CC pneumoniae is a common cause of respiratory disease in humans, and is  
CC also involved in the development of cardiovascular diseases such as

CC atherosclerosis, coronary artery disease, carotid artery stenosis,  
CC myocardial infarction, cerebrovascular disease, aortic aneurysm,  
CC claudication and stroke. The proteins and nucleic acids of the invention  
CC may be used in vaccines and pharmaceutical compositions for the  
CC prevention or treatment of chlamydial infections, particularly Chlamydia  
CC pneumoniae infections. The proteins may also be used in the detection of  
CC Chlamydia pneumoniae, and the nucleic acids may be used in PCR, branched  
CC DNA probe assay or blotting techniques for determining Chlamydia  
CC pneumoniae gene expression. The present sequence represents a  
CC specifically claimed DNA which encodes a Chlamydia pneumoniae protein of  
CC the invention. (Updated on 29-AUG-2003 to standardise OS field)  
XX  
SQ Sequence 669 BP; 227 A; 103 C; 136 G; 203 T; 0 U; 0 Other;  
  
Query Match 60.2%; Score 20; DB 6; Length 669;  
Best Local Similarity 57.1%; Pred. No. 1.1e+02;  
Matches 20; Conservative 5; Mismatches 10; Indels 0; Gaps 0;  
  
QY 2 CNCAYCGTGGCCNTCCNTAYAAAGYATYAGAR 36  
DB 513 CTCATGCATAGGCTTTACCATAGAAAGATAACGAG 479  
  
RESULT 62  
AAS36060/C  
ID AAS36060 standard; DNA; 5161 BP.  
XX  
AC AAS36060;  
DT 17-DEC-2001 (first entry)  
XX  
DE Human cardiovascular system antigen genomic DNA SEQ ID No 1560.  
XX  
KW Cardiovascular system antigen; human; mouse; rabbit; goat; horse; cat;  
KW chicken; sheep; immunosuppressive; antiarthritic; vasotropic; dog;  
KW antirheumatic; aniproliferative; cytostatic; cardiant; neuroprotective;  
KW cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;  
KW ophthalmological; vulnery; gene therapy; autoimmune disease; neoplasm;  
KW hyperproliferative disorder; breast; liver; cardiovascular disorder; ds;  
KW cerebrovascular disorder; nervous system disorder; bacterial infection;  
KW fungal infection; viral infection; ocular disorder; endocrine disorder;  
KW gastrointestinal disorder; renal disorder; respiratory disorder;  
KW wound healing; skin aging; organ transplantation; tissue regeneration;  
KW anti-infertility.  
XX  
OS Homo sapiens.  
XX  
XX WO20015321-A2.  
PN  
XX 02-AUG-2001.  
PD  
XX  
PF 17-JAN-2001; 2001WO-US0001340.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226686P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.

```
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0251989P.
PR 06-DEC-2000; 2000US-0251479P.
PR 06-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-451930/48.
XX
XX New cardiovascular system related polynucleotides and polypeptides,
XX useful for diagnosing, treating and/or preventing disorders of the
XX cardiovascular system.
XX
XX Claim 1; SEQ ID NO 1560; 674pp; English.
XX
XX Sequences AAS35741-AAS36942 represent genomic DNA molecules, which encode
XX the cardiovascular system antigen polypeptides of the invention.
XX Cardiovascular system antigens and their associated polynucleotides are
XX useful in the diagnosis, treatment and prevention of various types of
XX disorders in e.g. humans, mice, rabbits, goats, horses, cats, dogs,
XX chickens or sheep. A pathological condition can be determined by
XX detecting the presence or absence of a mutation in a cardiovascular
XX system antigen polynucleotide. The treatable disorders include autoimmune
XX diseases such as rheumatoid arthritis, hyperproliferative disorders such
XX as neoplasms of the breast or liver, cardiovascular disorders such as
XX cardiac arrest, cerebrovascular disorders such as cerebral ischaemia,
XX nervous system disorders such as Alzheimer's disease, infections caused
XX by bacteria, viruses and fungi, ocular disorders such as corneal
XX infection, endocrine disorders such as premature labour and infertility,
XX gastrointestinal disorders such as Crohn's disease, renal disorders such
XX as glomerulonephritis and respiratory disorders such as asthma and
XX pleurisy. The polypeptides can also be used to aid wound healing, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, to regenerate tissues and in chemotaxis. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
Query Match 60.2%; Score 20; DB 4; Length 5161;
Best Local Similarity 62.5%; Pred. No. 2e+02;
Matches 20; Conservative 4; Mismatches 8; Indels 0; Gaps 0;
QY 4 CAYGNTGGCCNTCCNTAYARGAYTAYGA 35
Db 4936 CATGCGTGGCGTCATATCTTAATGATTATGA 4905
RESULT 63
ADE46754/c
ID ADE46754 standard; DNA; 5161 BP.
XX
XX AC ADE46754;
XX
XX 29-JAN-2004 (first entry)
XX
XX Human cardiovascular system related genomic DNA #320.
XX
XX Human: cardiovascular system related polypeptide; cancer;
XX proliferative disorder; foetal abnormality; developmental abnormality;
XX haematopoietic disorder; AIDS; autoimmune disease; rheumatoid arthritis;
XX inflammation; allergy; neurological disorder; Alzheimer's disease;
XX Parkinson's disease; cognitive disorder; schizophrenia; asthma;
XX skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
XX cardiovascular disorder; angiogenic disorder; kidney disorder;
XX gastrointestinal disorder; pregnancy-related disorder;
XX endocrine disorder; gene; ds.
XX
XX Homo sapiens.
XX
XX OS
XX
XX PN US2003059908-A1.
XX
XX PD 27-MAR-2003.
XX
XX PF 07-MAR-2002; 2002US-00091504.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX 04-FEB-2000; 2000US-0180628P.
XX 24-FEB-2000; 2000US-0184664P.
XX 02-MAR-2000; 2000US-0186350P.
XX 16-MAR-2000; 2000US-0189874P.
XX 17-MAR-2000; 2000US-0190076P.
XX 18-APR-2000; 2000US-0198123P.
XX 19-MAY-2000; 2000US-0205515P.
XX 07-JUN-2000; 2000US-0209467P.
XX 28-JUN-2000; 2000US-0214886P.
XX 30-JUN-2000; 2000US-0215135P.
XX 07-JUL-2000; 2000US-0216647P.
XX 07-JUL-2000; 2000US-0216880P.
XX 11-JUL-2000; 2000US-0217487P.
XX 11-JUL-2000; 2000US-0217496P.
XX 14-JUL-2000; 2000US-0218290P.
XX 26-JUL-2000; 2000US-0220963P.
XX 26-JUL-2000; 2000US-0220964P.
XX 14-AUG-2000; 2000US-0224518P.
XX 14-AUG-2000; 2000US-0224519P.
XX 14-AUG-2000; 2000US-0225213P.
XX 14-AUG-2000; 2000US-0225214P.
XX 14-AUG-2000; 2000US-0225266P.
XX 14-AUG-2000; 2000US-0225267P.
XX 14-AUG-2000; 2000US-0225268P.
XX 14-AUG-2000; 2000US-0225270P.
XX 14-AUG-2000; 2000US-0225447P.
XX 14-AUG-2000; 2000US-0225757P.
XX 14-AUG-2000; 2000US-0225758P.
```

```
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226779P.
PR 22-AUG-2000; 2000US-0226861P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239335P.
PR 13-OCT-2000; 2000US-0239337P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0246471P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.

PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
PR 17-JAN-2001; 2001US-00764869.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM, Barash SC;
XX WPI; 2003-743766/70.
XX
XX New cardiovascular system related polynucleotides and polypeptides,
XX useful for preventing, treating, or ameliorating a medical condition,
XX such as cancer of cardiovascular tissues and cancer metastases.
XX
XX Claim 1; SEQ ID NO 1560; 262pp; English.
XX
XX The invention relates to human cardiovascular system related polypeptides
XX and the polynucleotides encoding them. The polypeptides, polynucleotides
XX and antibodies to the polypeptides are useful for diagnosing a
XX pathological condition or a susceptibility to a pathological condition,
XX for preventing, treating, or ameliorating a medical condition, such as
XX cancer of cardiovascular system tissues, proliferative disorders, foetal
XX and developmental abnormalities, haematopoietic disorders, diseases of
XX the immune system, AIDS, autoimmune diseases (e.g., rheumatoid
XX arthritis), inflammation, allergies, neurological disorders (e.g.,
XX Alzheimer's disease, Parkinson's disease), cognitive disorders,
XX schizophrenia, asthma, skin disorders (e.g., psoriasis), sepsis,
XX diabetes, atherosclerosis, cardiovascular disorders, angiogenic
XX disorders, kidney disorders, gastrointestinal disorders, pregnancy-
XX related disorders, endocrine disorders and infections. The nucleic acids
XX are also useful for chromosome identification, radiation hybrid mapping
XX or long-range restriction mapping. The polypeptides and polynucleotides
XX may also be used as food additives or preservatives to increase or
XX decrease storage capabilities, fat content or other nutritional
XX components. This sequence represents human cardiovascular system related
XX genomic DNA of the invention.
XX
XX Sequence 5161 BP; 1461 A; 1193 C; 1164 G; 1343 T; 0 U; 0 Other;
Query Match 60.2%; Score 20; DB 10; Length 5161;
Best Local Similarity 62.5%; Pred. No. 2e+02;
```









XX DE Plant full length insert polynucleotide seqid 33885.  
XX AC  
KW plant protectant; plant growth regulant; gene therapy; plant;  
KW recombinant DNA construct; physical array; plant breeding marker;  
KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;  
KW extreme osmotic condition; pathogen tolerance; pest tolerance;  
KW growth rate; cell cycle pathway; disease resistance;  
KW galactomannan production; lignin production; plant growth regulator;  
KW yield; plant growth; plant development; seed oil; protein yield;  
KW protein content; gene; ss.  
XX OS Unidentified.  
XX PN US2004034888-A1.  
XX PD 19-FEB-2004.  
XX PF 28-APR-2003; 2003US-00425114.  
XX PR 06-MAY-1999; 99US-00304517.  
XX PR 05-NOV-2001; 2001US-00985678.  
XX PA (LIUJ/) LIU J.  
PA (ZHOU/) ZHOU Y.  
PA (KOVA/) KOVALIC D K.  
PA (SCRE/) SCREEN S E.  
PA (TABA/) TABASKA J E.  
PA (CAOY/) CAO Y.  
XX PI Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JB, Cao Y;  
XX DR WPI; 2004-180133/17.  
XX PT New recombinant DNA construct, useful for improving plant tolerance to  
PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or  
PT pests, for conferring increased resistance to plant disease, or for  
PT improving yield.  
XX PS Claim 1; SEQ ID NO 33885; 15pp; English.  
XX CC The invention describes a recombinant DNA construct comprising a  
CC polynucleotide consisting of a sequence encoding an amino acid sequence  
CC available in electronic form from the US patent office at  
CC ftp.segdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide  
CC of the invention are also useful in physical arrays of molecules and as  
CC plant breeding markers. The recombinant DNA construct is useful for  
CC improving plant tolerance to cold, heat, drought, herbicides, extreme  
CC osmotic conditions, pathogens or pests, for manipulating growth rate in  
CC plant cells by modification of the cell cycle pathway, for conferring  
CC increased resistance to plant disease, for producing galactomannan,  
CC lignin or plant growth regulators, for increasing the rate of homologous  
CC recombination in plants, for improving yield by modification of  
CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake  
CC or by providing improved plant growth and development under at least one  
CC stress condition or for modifying seed oil or protein yield and/or  
CC content. This sequence represents a plant full length insert  
CC polynucleotide that can be used in the recombinant DNA construct of the  
CC invention.  
XX SQ Sequence 1083 BP; 267 A; 265 C; 321 G; 230 T; 0 U; 0 Other;  
Query Match 59.0%; Score 19.6; DB 13; Length 1083;  
Best Local Similarity 57.6%; Pred. No. 1.9e+02;  
Matches 19; Conservative 5; Mismatches 9; Indels 0; Gaps 0;  
QY 8 CNTGGCCNTCCNTAYARGTAYGARGTNA 40  
DB 46 CGTGGCCCGCCCGTCTAAGGACTACGAGTTCA 78  
RESULT 70  
ADX36401

ID ADX36401 standard; cDNA; 1305 BP.  
XX AC ADX36401;  
XX DT 21-APR-2005 (first entry)  
XX DE Plant full length insert polynucleotide seqid 19221.  
XX KW plant protectant; plant growth regulant; gene therapy; plant;  
KW recombinant DNA construct; physical array; plant breeding marker;  
KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;  
KW extreme osmotic condition; pathogen tolerance; pest tolerance;  
KW growth rate; cell cycle pathway; disease resistance;  
KW galactomannan production; lignin production; plant growth regulator;  
KW yield; plant growth; plant development; seed oil; protein yield;  
KW protein content; gene; ss.  
XX OS Unidentified.  
XX PN US2004034888-A1.  
XX PD 19-FEB-2004.  
XX PF 28-APR-2003; 2003US-00425114.  
XX PR 06-MAY-1999; 99US-00304517.  
XX PR 05-NOV-2001; 2001US-00985678.  
XX PA (LIUJ/) LIU J.  
PA (ZHOU/) ZHOU Y.  
PA (KOVA/) KOVALIC D K.  
PA (SCRE/) SCREEN S E.  
PA (TABA/) TABASKA J E.  
PA (CAOY/) CAO Y.  
XX PI Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JB, Cao Y;  
XX DR WPI; 2004-180133/17.  
XX PT New recombinant DNA construct, useful for improving plant tolerance to  
PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or  
PT pests, for conferring increased resistance to plant disease, or for  
PT improving yield.  
XX PS Claim 1; SEQ ID NO 19221; 15pp; English.  
XX CC The invention describes a recombinant DNA construct comprising a  
CC polynucleotide consisting of a sequence encoding an amino acid sequence  
CC available in electronic form from the US patent office at  
CC ftp.segdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide  
CC of the invention are also useful in physical arrays of molecules and as  
CC plant breeding markers. The recombinant DNA construct is useful for  
CC improving plant tolerance to cold, heat, drought, herbicides, extreme  
CC osmotic conditions, pathogens or pests, for manipulating growth rate in  
CC plant cells by modification of the cell cycle pathway, for conferring  
CC increased resistance to plant disease, for producing galactomannan,  
CC lignin or plant growth regulators, for increasing the rate of homologous  
CC recombination in plants, for improving yield by modification of  
CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake  
CC or by providing improved plant growth and development under at least one  
CC stress condition or for modifying seed oil or protein yield and/or  
CC content. This sequence represents a plant full length insert  
CC polynucleotide that can be used in the recombinant DNA construct of the  
CC invention.  
XX SQ Sequence 1305 BP; 302 A; 350 C; 368 G; 285 T; 0 U; 0 Other;  
Query Match 59.0%; Score 19.6; DB 13; Length 1305;  
Best Local Similarity 57.6%; Pred. No. 2e+02;  
Matches 19; Conservative 5; Mismatches 9; Indels 0; Gaps 0;  
QY 8 CNTGGCCNTCCNTAYARGTAYGARGTNA 40

Db 235 CGTCGCCGCCCGCTCTAAGGACTACGAGTTCA 267

RESULT 71

ABZ90108

ID ABZ90108 standard; DNA; 286 BP.

XX AC ABZ90108;

XX DT 17-OCT-2003 (first entry)

XX DE Human nucleic acid sequence.

XX KW Human; antisense; lung dysfunction; nasal airway dysfunction; anti-inflammatory steroid; ubiquinone; anti-inflammatory; antiallergic; antiasthmatic; hypotensive; immunosuppressive; cytosolic; gene therapy; antisense gene therapy; respiratory; lung; adenosine sensitivity; adenosine receptor; bronchodilation; bronchoconstriction; lung allergy; lung inflammation; respiratory disease; ds.

XX OS Homo sapiens.

XX PN WO200285308-A2.

XX PD 31-OCT-2002.

XX PF 23-APR-2002; 2002WO-US013135.

XX PR 24-APR-2001; 2001US-0286137P.

XX PA (EPIG-) EPIGENESIS PHARM INC.

XX PI Nyce JW, Li Y, Sandrasagra A, Katz E, Pabalan J, Aguilar D;

XX PI Miller S, Tang L, Shahabuddin S;

XX DR WPI; 2003-229219/22.

XX PT Pharmaceutical composition for treating ailments associated with impaired

XX PT respiration, has oligo(s) antisense to specific gene(s) or its

XX PT corresponding RNAs, and glucocorticoid or non-glucocorticoid steroid or

XX PT ubiquinone.

XX PS Disclosure; SEQ ID NO 5350; 872pp; English.

XX CC The invention relates to a novel pharmaceutical composition, which has a

XX CC first active agent comprising an oligonucleotide antisense to the

XX CC initiation codon, coding region, 5' or 3' end genomic flanking regions,

XX CC 5' and 3' intron-exon junctions, or regions within 2-10 nucleotides of

XX CC junctions of genes encoding a polypeptide associated with lung and/or

XX CC nasal airway dysfunction and a second active agent comprising an

XX CC anti-inflammatory steroid and ubiquinone. A composition of the invention

XX CC has anti-inflammatory, antiallergic, antiasthmatic, hypotensive,

XX CC immunosuppressive, and cytosolic activity. The composition may have a

XX CC use in antisense gene therapy. The composition is useful for treating or

XX CC preventing a respiratory, lung or malignant disease or condition, also

XX CC for enhancing the prophylactic or therapeutic respiratory effect of an

XX CC anti-inflammatory steroid in a subject, for reducing or depleting levels

XX CC of, or reducing sensitivity to adenosine, reducing levels of adenosine

XX CC receptor, producing bronchodilation, increasing levels of ubiquinone or

XX CC lung surfactant in a subject's tissue, or treating bronchoconstriction,

XX CC lung inflammation, lung allergies, or a respiratory disease or condition.

XX CC Note: The sequence data for this patent is not represented in the printed

XX CC specification, but was obtained in electronic format directly from WIPO

XX CC at ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 286 BP; 88 A; 65 C; 61 G; 72 T; 0 U; 0 Other;

Query Match 58.48; Score 19.4; DB 10; Length 286;

Best Local Similarity 48.87; Pred. No. 1.6e+02;

Matches 20; Conservative 7; Mismatches 14; Indels 0; Gaps 0;

QY 2 CNCAYGNTGCCGTCNCCNCTAAGGACTACGAGTTCAAR 42

|||:||||| |||:||||| |||:||||| |||:|||||

Db 92 CACCTGCTGATATCTCTACTACAAACATTTGCGGTCAA 132

RESULT 72

ABD26338

ID ABD26338 standard; DNA; 286 BP.

XX AC ABD26338;

XX DT 29-JUL-2004 (first entry)

XX DE AA459692 DNA fragment.

XX KW Human; antisense; bronchoconstriction; allergy; hyposecretion; pain; respiratory tract inflammation; adenosine sensitivity; lung; cancer; surfactant depletion; antiallergic; anti-inflammatory; antiasthmatic; analgesic; hypotensive; immunosuppressive; cytosolic; cystic fibrosis; beta-adrenergic agonist; respiratory disease; pulmonary vasoconstriction; respiratory distress syndrome; allergic rhinitis; pulmonary hypertension; emphysema; chronic obstructive pulmonary disease; cancer; bronchitis; pulmonary transplantation rejection; ds.

XX OS Homo sapiens.

XX PN WO200285309-A2.

XX PD 31-OCT-2002.

XX PF 23-APR-2002; 2002WO-US013143.

XX PR 24-APR-2001; 2001US-0286036P.

XX PA (EPIG-) EPIGENESIS PHARM INC.

XX PI Nyce JW, Li Y, Sandrasagra A, Katz E, Pabalan J, Aguilar D;

XX PI Miller S, Tang L, Shahabuddin S;

XX DR WPI; 2003-093058/08.

XX PT Pharmaceutical composition for treating asthma, has antisense

XX PT oligonucleotide containing less percentage of adenosine, targeted to

XX PT nucleic acids associated with lung airway or lung dysfunction, and

XX PT bronchodilating agent.

XX PS Claim 15; SEQ ID NO 5350; 763pp; English.

XX CC This invention describes a novel composition (a) a first active agent,

XX CC comprising oligonucleotides, effective for alleviating

XX CC bronchoconstriction, respiratory tract inflammation, allergies and

XX CC reducing adenosine sensitivity, levels of adenosine (A) or (A) receptors,

XX CC surfactant depletion or hyposecretion, when administered to a mammal. The

XX CC oligonucleotides are derived from a gene encoding or regulating

XX CC expression of a target polypeptide associated with lung airway or lung

XX CC dysfunction or cancer and can be anti-sense to the corresponding mRNA.

XX CC The invention also describes a kit, that comprises: (a) a delivery

XX CC device, in separate containers, (b) the oligonucleotides, (c)

XX CC instructions for adding a carrier and for use of the kit. The composition

XX CC of the invention has antiallergic, anti-inflammatory, antiasthmatic,

XX CC analgesic, hypotensive, immunosuppressive and cytostatic activity, is a

XX CC beta-adrenergic agonist. The composition is useful for preventing or

XX CC treating a respiratory, lung or malignant disease. The administered

XX CC composition comprises oligo and is administered to reduce the production

XX CC or availability, or to increase the degradation of the target mRNA or to

XX CC reduce the amount of target polypeptide present in the lungs. The

XX CC pulmonary obstruction, and/or bronchoconstriction and/or lung

XX CC inflammation, allergies and/or surfactant hypoproduction are associated

XX CC with a disease or condition such as pulmonary vasoconstriction,

XX CC inflammation, allergies, asthma, impeded respiration, respiratory

XX CC distress syndrome, pain, cystic fibrosis, allergic rhinitis, pulmonary

XX CC hypertension, emphysema, chronic obstructive pulmonary disease, pulmonary

XX CC transplantation rejection, pulmonary infections, bronchitis or cancer.

XX CC The reduced adenosine content of the anti-sense oligos corresponding to

XX CC thymidines present in the target RNA serves to prevent the breakdown of





RESULT 76	
ADW28597/c	
ID	ADW28597 standard; cDNA; 918 BP.
XX	
AC	ADW28597;
XX	
DT	10-MAR-2005 (first entry)
XX	
DE	cDNA encoding human formamido pyrimidine-DNA glycosylase 11.77.
XX	
KW	Formamido pyrimidine-DNA glycosylase 11.77; cancer; neoplasm;
KW	HIV infection; infection; immune disorder; cytostatic; anti-HIV; gene;
KW	ss.
XX	
OS	Homo sapiens.
XX	
Key	Location/Qualifiers
FT	CDS
FT	287..610
FT	/*tag= a
FT	/product= "Formamido pyrimidine-DNA glycosylase 11.77"
XX	
PN	CN1519320-A.
XX	
PD	11-AUG-2004.
XX	
PF	21-JAN-2003; 2003CN-00115077.
XX	
PR	21-JAN-2003; 2003CN-00115077.
XX	
PA	(BIOW-) BIO WINDOW GENE DEV INC.
XX	
PI	Mao Y, Xie Y;
XX	
WPI	WPI; 2004-785373/78.
DR	P-PDB; ADW28598.
XX	
PT	New human formamido pyrimidine-glycosylase 11.77 and encoding
PT	polynucleotide for treating diseases such as cancer, HIV infection, and
PT	immunopathy.
XX	
PS	Claim 6; SEQ ID NO 1; 22pp; Chinese.
XX	
CC	The invention relates to the isolation of human formamido pyrimidine-DNA
CC	glycosylase 11.77, and the polynucleotide sequence encoding it. Also
CC	described is the preparation of the protein by DNA recombination, the
CC	application of the polypeptide in treating diseases (e.g. cancer, HIV
CC	infection, and immunopathy), the antagonist of the polypeptide and its
CC	medical action, and the application of the polynucleotide. This sequence
CC	encodes human formamido pyrimidine-DNA glycosylase 11.77.
XX	
SQ	Sequence 918 BP; 164 A; 271 C; 271 G; 212 T; 0 U; 0 Other;
Query Match	58.4%; Score 19.4; DB 13; Length 918;
Best Local Similarity	55.6%; Pred. No. 2.3e+02;
Matches	20; Conservative 5; Mismatches 11; Indels 0; Gaps 0;
QY	1 GCNCAYGCGTGGCCNTCNCNTAYAAAGAYTAGAR 36
Db	482 GCACAGCATGGCCTTGGCCAGACACAAACCAAGAG 447
RESULT 77	
ADT04975/c	
ID	ADT04975 standard; DNA; 992 BP.
XX	
AC	ADT04975;
XX	
DT	02-DEC-2004 (first entry)
XX	
DE	Haemophilus influenzae (NTHI) contig DNA sequence - SEQ ID 11.
XX	
KW	middle ear bacterial infection; nasopharynx bacterial infection; ds;
KW	contig.



```

XX Haemophilus influenzae.
OS
XX WO2004078949-A2.
PN
XX
XX 16-SEP-2004.
PD
XX
XX 05-MAR-2004; 2004WO-US007001.
PF
XX
XX 06-MAR-2003; 2003US-0453134P.
PR
XX
XX (CHIL-) CHILDRENS HOSPITAL INC.
PA
XX
XX Bakaletz LO, Munson RS, Dyer DW;
PI
XX
XX WPI; 2004-662422/64.
DR
XX
XX New polynucleotides of nontypeable strain of Haemophilus influenzae,
PT
XX useful for treating or preventing NTHi bacterial infections of the middle
PT
XX ear and/or nasopharynx.
PT
XX
XX Example 1; SEQ ID NO 11; 88pp; English.
PS
XX
XX The invention comprises nucleotide sequences (genes) from the genome of a
CC
XX nontypeable strain of Haemophilus influenzae (NTHi). The NTHi DNA
CC
XX sequences of the invention are useful for treating or preventing NTHi
CC
XX bacterial infections of the middle ear and/or nasopharynx. The present
CC
XX nucleic acid represents an NTHi contig sequence of the invention.
CC
XX
XX Sequence 992 BP; 321 A; 173 C; 210 G; 287 T; 0 U; 1 Other;
SQ
XX
XX Query Match 58.4%; Score 19.4; DB 13; Length 992;
XX Best Local Similarity 48.8%; Pred. No. 2.3e+02;
XX Matches 20; Conservative 7; Mismatches 14; Indels 0; Gaps 0;
XX
XX QY 2 CNCAYGCGTGGCCNTCCNTAYAAAGAYTAYGARGTNAAR 42
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 939 CGCTTGATGGACATCCCTCCCAATATGTTTACGAAGTANAA 899
XX
XX
XX RESULT 78
XX AAH99671
XX ID AAH99671 standard; cDNA; 1757 BP.
XX
XX AAH99671;
AC
XX
XX 16-OCT-2001 (first entry)
DT
XX
XX Human protein encoding cDNA sequence SEQ ID NO:506.
DE
XX
XX Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
KW
XX anti-inflammatory; antirheumatic; antiarthritic; immunosuppressive;
KW
XX antibacterial; endocrine; cardiant; central nervous system; virucide;
KW
XX anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
KW
XX antiaggregant; haemostatic; vulnerary; antiulcer; osteopathic; eczema;
KW
XX dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;
KW
XX neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
KW
XX immunostimulant; gene therapy; antitense therapy; vaccine; inflammation;
KW
XX antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
KW
XX cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
KW
XX genetic disease; haematopoietic disorder; platelet disorder; asthma;
KW
XX thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
KW
XX allergic rhinitis; diabetes; multiple sclerosis; depression;
KW
XX Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
KW
XX neurological disorder; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200153455-A2.
PN
XX
XX 26-JUL-2001.
PD
XX
XX 22-DEC-2000; 2000WO-US035017.
PF

```

```

XX 23-DEC-1999; 99US-00471275.
PR
XX 21-JAN-2000; 2000US-00488725.
PR
XX 25-APR-2000; 2000US-00552317.
PR
XX (HYSE-) HYSEQ INC.
PA
XX
XX Tang YT, Liu C, Drmanac RT;
PI
XX
XX WPI; 2001-457603/49.
DR
XX P-PSDB; AAM25730.
DR
XX
XX Isolated human polynucleotides encoding polypeptides, useful for the
PT
XX treatment and diagnosis of e.g. cancer, ulcers and HIV infection.
PT
XX
XX Claim 1; Page 569; 1217pp; English.
PS
XX
XX AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
CC
XX AAM25963. The proteins can have activities based on the tissues and cells
CC
XX they are expressed in, such as: antiinflammatory; antirheumatic;
CC
XX antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;
CC
XX central nervous system; virucide; anti-HIV; fungicide; antimutagen;
CC
XX cardiovascular; antianaemic; antiaggregant; haemostatic; vulnerary;
CC
XX antidiabetic; cytostatic; neuroprotective; antiallergic; antiasthmatic;
CC
XX antiparkinsonian; and immunostimulant. The proteins and polynucleotides
CC
XX encoding them can be used in gene therapy, antitense therapy and vaccine
CC
XX production. The proteins and polynucleotides are useful for screening for
CC
XX agonists or antagonists of a protein and for the treatment and diagnosis
CC
XX of disorders associated with the activity of a protein e.g. inflammation,
CC
XX rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
CC
XX neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
CC
XX infections, autoimmunity, genetic diseases, haematopoietic disorders,
CC
XX anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
CC
XX osteoporosis, severe combined immunodeficiency, eczema, allergic
CC
XX rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
CC
XX Alzheimer's disease, Parkinson's disease, neurodegenerative and
CC
XX neurological disorders
XX
XX SQ Sequence 1757 BP; 401 A; 526 C; 464 G; 366 T; 0 U; 0 Other;
XX
XX Query Match 58.4%; Score 19.4; DB 4; Length 1757;
XX Best Local Similarity 55.6%; Pred. No. 2.7e+02;
XX Matches 20; Conservative 5; Mismatches 11; Indels 0; Gaps 0;
XX
XX QY 1 GCNCAYGCGTGGCCNTCCNTAYAAAGAYTAYGAR 36
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 452 GCACACGCGATGGCTTGGCCAGACACAAACCAAGAG 487
XX
XX
XX RESULT 79
XX ABX71365/c
XX ID ABX71365 standard; cDNA; 2321 BP.
XX
XX ABX71365;
AC
XX
XX 14-APR-2003 (first entry)
DT
XX
XX Human signal transduction-associated cDNA from clone DKF2phtes3_21d4.
DE
XX
XX Human; gene; gene therapy; vaccine; disease treatment; detection; ss.
KW
XX
XX Homo sapiens.
OS
XX
XX WO200112659-A2.
PN
XX
XX 22-FEB-2001.
PD
XX
XX 18-AUG-2000; 2000WO-1B001496.
PF
XX
XX 18-AUG-1999; 99US-0149499P.
PR
XX 28-SEP-1999; 99US-0156503P.
PR
XX

```

PA (GEHU-) GERMAN HUMAN GENOME PROJECT.  
XX Wiemann S;  
PI WPI; 2001-327840/34.  
DR P-PSDB; ABUS3098.  
XX  
XX Nucleic acids having the sequences of clones isolated from libraries of  
PT different human tissues, useful in recombinant DNA methodologies.  
XX  
XX Claim 1; Page 705; 1095pp; English.  
XX  
XX This invention describes novel polynucleotides and polypeptides isolated  
CC from human cDNA libraries which can be used for gene therapy or in  
CC vaccines. The polynucleotides of the invention and antibodies encoded by  
CC them may be used in the prevention, diagnosis and treatment of diseases  
CC associated with inappropriate polypeptide expression. The products of the  
CC invention may also be used to identify modulators of expression and  
CC activity and to down regulate expression and activity. The antibodies of  
CC the invention may also be used as diagnostic agents for detecting the  
CC presence of polypeptides in samples. This sequence encodes a polypeptide  
CC described in the disclosure of the invention  
XX  
SQ Sequence 2321 BP; 461 A; 626 C; 739 G; 495 T; 0 U; 0 Other;  
Query Match 58.4%; Score 19.4; DB 5; Length 2321;  
Best Local Similarity 55.6%; Pred. No. 2.9e+02;  
Matches 20; Conservative 5; Mismatches 11; Indels 0; Gaps 0;  
QY 1 GCNCGCAGTGGCCNTCCNTAYAAAGAYTAYGAR 36  
Db 1862 GCACAGCATGGCTTGGCCAGACACAAACCAAGAG 1827  
RESULT 80  
ABZ11510/c  
ID ABZ11510 standard; cDNA; 2416 BP.  
XX  
AC ABZ11510;  
XX  
DT 20-JAN-2003 (first entry)  
XX  
DE Human polynucleotide SEQ ID NO 392.  
XX  
KW Human; genome mapping; gene therapy; food supplement; virus; fungus;  
KW cell-proliferative disorder; neurodegenerative disease; bacterial;  
KW Parkinson's disease; Alzheimer's disease; autoimmune disease;  
KW multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;  
KW arthritis; cytostatic; immunomodulator; nontropic; neuroprotective;  
KW antiparkinsonian; antidiabetic; immunosuppressive; dermatological;  
KW haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide;  
KW antiarthritic; gene; ss.  
XX  
OS Homo sapiens.  
XX  
XX WO200270539-A2.  
PN  
XX 12-SEP-2002.  
XX  
PF 05-MAR-2002; 2002WO-US005095.  
XX  
PR 05-MAR-2001; 2001US-00799451.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;  
XX Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;  
PI Wehrman T, Wang J, Wang D, Drmanac RT;  
XX WPI; 2002-759812/82.  
DR  
XX P-PSDB; ABP69293.  
XX  
XX New polynucleotides comprising sequences assembled from expressed

PT sequence tags (ESTs), useful for treating cell-proliferative,  
PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet  
PT or coagulation disorders.  
XX  
XX Claim 1; SEQ ID NO 392; 1012pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated polynucleotide (I) comprising a  
CC nucleotide sequence selected from any of 948 sequences (ABZ11119-  
CC ABZ12066) or their mature protein coding portion, active domain coding  
CC protein or complementary sequences. The polynucleotides are useful for  
CC identifying expressed genes or for physical mapping of human genome. The  
CC encoded polypeptides (ABP68902-ABP69849) are useful as molecular weight  
CC markers, as a food supplement, for generating antibodies, in medical  
CC imaging, screening and diagnostic assays and for treating cell-  
CC proliferative disorders (cancer), neurodegenerative diseases (Parkinson's  
CC or Alzheimer's disease), autoimmune diseases (multiple sclerosis,  
CC diabetes, lupus) genetic disorders, myeloid or lymphoid disorders,  
CC platelet or coagulation disorders, wound, burns, incision, ulcers, liver  
CC or lung fibrosis, infections (bacterial, viral, fungal, parasitic),  
CC arthritis, etc. Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 2416 BP; 457 A; 634 C; 748 G; 577 T; 0 U; 0 Other;  
Query Match 58.4%; Score 19.4; DB 6; Length 2416;  
Best Local Similarity 55.6%; Pred. No. 3e+02;  
Matches 20; Conservative 5; Mismatches 11; Indels 0; Gaps 0;  
QY 1 GCNCGCAGTGGCCNTCCNTAYAAAGAYTAYGAR 36  
Db 1965 GCACAGCATGGCTTGGCCAGACACAAACCAAGAG 1930  
RESULT 81  
ADM44028/c  
ID ADM44028 standard; cDNA; 2416 BP.  
XX  
AC ADM44028;  
XX  
DT 03-JUN-2004 (first entry)  
XX  
DE Novel human arginine-rich protein cDNA #392.  
XX  
KW ss; gene; human; arginine-rich protein; cancer; inflammation;  
KW genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN US2004053250-A1.  
XX  
PD 19-MAR-2004.  
XX  
XX 21-NOV-2002; 2002US-00302172.  
XX  
XX 05-MAR-2001; 2001US-00799451.  
PR  
XX 05-MAR-2002; 2002WO-US005095.  
PR  
XX 20-AUG-2002; 2002US-00225251.  
XX  
XX (TANG/) TANG Y T.  
XX (XUEA/) XUE A.  
XX (DRMA/) DRMANAC R T.  
XX  
XX Tang YT, Xue A, Drmanac RT;  
PI  
XX WPI; 2004-238579/22.  
DR  
XX New isolated arginine-rich protein-like polynucleotides and polypeptides,  
PT useful for diagnosing and/or treating conditions associated with aberrant  
PT activity of the arginine-rich polypeptides, such as cancer and  
PT inflammation.  
XX  
XX Disclosure; SEQ ID NO 392; 51pp; English.

XX The invention relates to an isolated polynucleotide. The methods and  
CC compositions of the present invention are useful for the diagnosis and/or  
CC treatment of diseases or conditions associated with aberrant expression  
CC or activity of the arginine-rich protein-like polypeptides, such as  
CC cancer and inflammation. They can also be used in forensics, gene  
CC mapping, identification of mutations responsible for genetic disorders,  
CC and in assessing biodiversity. The present sequence represents a novel  
CC human arginine-rich protein cDNA.

XX Sequence 2416 BP; 457 A; 634 C; 748 G; 577 T; 0 U; 0 Other;  
SQ

Query Match 58.4%; Score 19.4; DB 12; Length 2416;  
Best Local Similarity 55.6%; Pred. No. 3e+02;  
Matches 20; Conservative 5; Mismatches 11; Indels 0; Gaps 0;

QY 1 GCNCAVCNTGGCCTTCCGAGCAACCAAGAG 36  
Db 1965 GCACACGATGGCTTGGCCAGACACAAACCAAGAG 1930

RESULT 82  
AAC47167  
ID AAC47167 standard; DNA; 3688 BP.

XX  
AC AAC47167;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 52804.

XX Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway; metabolic pathway;  
KW promoter; termination sequence; ss.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-00301439.  
XX  
PR 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 04-MAY-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.

21-MAY-1999; 99US-0135353P.  
24-MAY-1999; 99US-0135629P.  
25-MAY-1999; 99US-0136021P.  
27-MAY-1999; 99US-0136392P.  
28-MAY-1999; 99US-0136782P.  
01-JUN-1999; 99US-0137222P.  
03-JUN-1999; 99US-0137528P.  
04-JUN-1999; 99US-0137502P.  
07-JUN-1999; 99US-0137724P.  
08-JUN-1999; 99US-0138094P.  
10-JUN-1999; 99US-0138540P.  
14-JUN-1999; 99US-0138847P.  
16-JUN-1999; 99US-0139119P.  
16-JUN-1999; 99US-0139452P.  
16-JUN-1999; 99US-0139453P.  
17-JUN-1999; 99US-0139492P.  
18-JUN-1999; 99US-0139454P.  
18-JUN-1999; 99US-0139455P.  
18-JUN-1999; 99US-0139456P.  
18-JUN-1999; 99US-0139457P.  
18-JUN-1999; 99US-0139458P.  
18-JUN-1999; 99US-0139459P.  
18-JUN-1999; 99US-0139460P.  
18-JUN-1999; 99US-0139461P.  
18-JUN-1999; 99US-0139462P.  
18-JUN-1999; 99US-0139463P.  
18-JUN-1999; 99US-0139750P.  
18-JUN-1999; 99US-0139763P.  
21-JUN-1999; 99US-0139817P.  
22-JUN-1999; 99US-0139899P.  
23-JUN-1999; 99US-0140353P.  
23-JUN-1999; 99US-0140354P.  
24-JUN-1999; 99US-0140695P.  
28-JUN-1999; 99US-0140823P.  
29-JUN-1999; 99US-0140991P.  
30-JUN-1999; 99US-0141287P.  
01-JUL-1999; 99US-0141842P.  
01-JUL-1999; 99US-0142154P.  
02-JUL-1999; 99US-0142055P.  
06-JUL-1999; 99US-0142390P.  
08-JUL-1999; 99US-0142803P.  
09-JUL-1999; 99US-0142920P.  
12-JUL-1999; 99US-0142977P.  
13-JUL-1999; 99US-0143542P.  
14-JUL-1999; 99US-0143624P.  
15-JUL-1999; 99US-0144005P.  
16-JUL-1999; 99US-0144085P.  
16-JUL-1999; 99US-0144086P.  
19-JUL-1999; 99US-0144325P.  
19-JUL-1999; 99US-0144331P.  
19-JUL-1999; 99US-0144332P.  
19-JUL-1999; 99US-0144333P.  
19-JUL-1999; 99US-0144334P.  
20-JUL-1999; 99US-0144335P.  
20-JUL-1999; 99US-0144352P.  
20-JUL-1999; 99US-0144632P.  
20-JUL-1999; 99US-0144884P.  
21-JUL-1999; 99US-0144814P.  
21-JUL-1999; 99US-0145086P.  
21-JUL-1999; 99US-0145088P.  
22-JUL-1999; 99US-0145085P.  
22-JUL-1999; 99US-0145087P.  
22-JUL-1999; 99US-0145089P.  
22-JUL-1999; 99US-0145192P.  
23-JUL-1999; 99US-0145145P.  
23-JUL-1999; 99US-0145218P.  
23-JUL-1999; 99US-0145224P.  
26-JUL-1999; 99US-0145226P.  
27-JUL-1999; 99US-0145313P.  
27-JUL-1999; 99US-0145918P.  
27-JUL-1999; 99US-0145919P.  
28-JUL-1999; 99US-0145951P.  
02-AUG-1999; 99US-0146386P.

```
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 01-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0156559P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.

PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 58.4%; Score 19.4; DB 3; Length 3688;
Best Local Similarity 55.6%; Pred. No. 3.4e+02;
Matches 20; Conservative 5; Mismatches 11; Indels 0; Gaps 0;

QY 7 GCNTGGCCNTCCNTAYAARGAYTAYGARGTNAAR 42
Db 1240 GCTGGACCAACACCGTATACAGACTACGATGCAAG 1275

RESULT 83
ADFO1135
ID ADF01135 standard; DNA; 3738 BP.
XX
AC ADF01135;
XX
DT 12-FEB-2004 (first entry)
XX
DE Bacterial polynucleotide #1420.
XX
KW Proteus mirabilis infection; bacterial infection; antibacterial;
immunostimulant; gene; ds.
XX
OS Proteus mirabilis.
XX
PN US6605709-B1.
XX
PD 12-AUG-2003.
XX
PF 05-APR-2000; 2000US-00543681.
PR 09-APR-1999; 99US-0128706P.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
PA Breton GL;
PI
XX
XX WPI: 2003-895291/82.
DR P-PSDB; ADF05307.
XX
XX New Proteus mirabilis polypeptides and polynucleotides, useful as
reagents for diagnosis of bacterial disease, as components of
antibacterial vaccines, as targets for antibacterial drugs, or as
biocontrol agents for plants.
XX
XX Disclosure; SEQ ID NO 1420; 870pp; English.
XX
XX The invention relates to new Proteus mirabilis polypeptides and
polynucleotides. The invention also relates to antibodies against the
polypeptides, methods for producing the polypeptides, a method of
generating vaccines for immunising an individual against P. mirabilis, a
method for evaluating a compound for the ability to bind a P. mirabilis
polypeptide and a method for screening test compounds for anti-bacterial
activity. The polypeptides and polynucleotides are useful as molecular
targets for diagnosing, preventing and treating pathological conditions
resulting from bacterial infection, as reagents for diagnosis of
bacterial diseases, as components of antibacterial vaccines, as targets
for antibacterial drugs or as bio-control agents for plants. This
sequence represents a Proteus mirabilis polynucleotide of the invention.
XX
SQ Sequence 3738 BP; 1241 A; 730 C; 722 G; 1045 T; 0 U; 0 Other;

Query Match 58.4%; Score 19.4; DB 10; Length 3738;
Best Local Similarity 52.6%; Pred. No. 3.4e+02;
Matches 20; Conservative 6; Mismatches 12; Indels 0; Gaps 0;

QY 5 AVCGNTGGCCNTCCNTAYAARGAYTAYGARGTNAAR 42
Db 3278 AGCATGGATTGCACACTTCAAGATTACAAGTGAA 3315
```

```

RESULT 84
AAZ11190/c
ID AAZ11190 standard; cDNA; 418 BP.
XX
AC AAZ11190;
XX
AT 04-NOV-1999 (first entry)
XX
DE Interleukin-3 coding sequence from b2HFLS20W cDNA library.
XX
KW Interleukin-3; IL-3; human; bone marrow; blood cell; immunodeficiency;
KW infection; immune disorder; rheumatoid arthritis; chemotherapy; therapy;
KW myeloid cell activation; lymphoid cell; leukopaenia; dendritic-like cell;
KW haematopoietic precursor cell; vaccine development; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 2..202
FT /tag= a
FT /product= "Interleukin-3"
XX
PN WO9941382-A2.
XX
PD 19-AUG-1999.
XX
PF 17-FEB-1999; 99WO-US001484.
XX
PR 17-FEB-1998; 98US-00024820.
XX
PR 23-OCT-1998; 98US-00177467.
XX
PR 11-DEC-1998; 98US-00209534.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
PI Leshkowitz D, Kita D, Ford J;
XX
DR WPI; 1999-527369/44.
DR P-PSDB; AAY32936.
XX
PT New Interleukin-3 useful for treating conditions requiring increases in
PT numbers or differentiation of hematopoietic cells, e.g. autoimmune
PT diseases, leukopenia.
XX
PS Claim 1; Fig 2; 158pp; English.
XX
CC This sequence encodes the novel interleukin-3 (IL-3) of the invention.
CC This sequence was isolated from the human foetal liver-spleen tissue cDNA
CC library b2HFLS20W. The IL-3 polypeptides are useful in vitro for
CC stimulating growth and activation of bone marrow and blood cells before
CC infusion of these cells into patients. They can be used to treat various
CC immunodeficiencies (e.g. resulting from viral infection) or immune
CC disorders (e.g. rheumatoid arthritis) or other conditions in which
CC haematopoietic cell populations have been reduced/destroyed e.g. due to
CC radiation or chemotherapy. They are especially useful to activate mature
CC myeloid and/or lymphoid cells and especially to treat leukopaenia. The
CC polypeptides can also be used to stimulate differentiation in
CC haematopoietic precursor cells, especially haematopoietic stem or
CC progenitor cells, lymphoid cells or myeloid (especially monocyte) cells,
CC e.g. the differentiation of haematopoietic precursor cells into dendritic-
CC like cells. This is useful in vivo to increase the number of dendritic-
CC like cells to enhance the immune response against infections and diseases
CC e.g. to treat pathogenic infections, autoimmune diseases and in
CC transplantation; in vitro production of dendritic cells is useful
CC clinically (e.g. in vaccine development) since isolation of these cells
CC from tissues is difficult. IL-3 can be combined with a carrier in
CC compositions useful as above, and used to generate antibodies useful in
CC assays and therapeutically. The polynucleotides are useful e.g. as
CC hybridisation probes, as oligomers for PCR, for chromosome and gene
CC mapping and for polypeptide production
XX
SQ Sequence 418 BP; 97 A; 111 C; 109 G; 101 T; 0 U; 0 Other;
Query Match 57.8%; Score 19.2; DB 2; Length 418;
Best Local Similarity 53.8%; Pred. No. 2.2e+02;
Matches 21; Conservative 5; Mismatches 13; Indels 0; Gaps 0;
QY 4 CAYGCNTGGCNCNTCCNTAYAAAGAYTAYGARGTNAAR 42
DB 82 CATGCTGGTCTCTCCCTAAGAGGATTAGGAAATAAAG 44
RESULT 85
AAD39735/c
ID AAD39735 standard; DNA; 418 BP.
XX
AC AAD39735;
XX
AT 22-OCT-2002 (first entry)
XX
DE Human interleukin-3 (IL-3) DNA #1.
XX
KW Human; interleukin-3; IL-3; gene mapping; chromosome mapping; radiation;
KW haematopoiesis; chemotherapy; pathogen infection; autoimmune disease;
KW transplant rejection; aplastic anaemia; leukopaenia; immunosuppressive;
KW leukaemia; cytostatic; immunostimulant; haemostatic; ds.
XX
OS Homo sapiens.
XX
PN US2002058018-A1.
XX
PD 16-MAY-2002.
XX
PF 23-FEB-2001; 2001US-00792246.
XX
PR 17-AUG-1999; 99US-00376732.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Ford J;
XX
DR WPI; 2002-499505/53.
XX
PT New nucleic acid encoding a human interleukin-3, useful for promoting
PT growth and differentiation of hematopoietic precursor cells, e.g. for
PT treating leucopenia.
XX
PS Claim 1; Fig 2; 58pp; English.
XX
CC The present invention relates to human interleukin-3 (IL-3) proteins and
CC polynucleotides encoding such proteins. Sequences of the invention are
CC useful for gene mapping, chromosome mapping and to produce antisense
CC sequences. They are used to promote survival, growth and differentiation
CC of haematopoietic precursor cells, for the treatment of any condition
CC associated with inadequate haematopoiesis, e.g. side effects of radiation
CC or chemotherapy, pathogen infections, autoimmune diseases, transplant
CC rejection, aplastic anaemia, leukaemia, leukopaenia and in screening for
CC specific agonists or antagonists, potential therapeutic agents. The
CC present sequence is human IL-3 DNA
XX
SQ Sequence 418 BP; 97 A; 111 C; 109 G; 101 T; 0 U; 0 Other;
Query Match 57.8%; Score 19.2; DB 6; Length 418;
Best Local Similarity 53.8%; Pred. No. 2.2e+02;
Matches 21; Conservative 5; Mismatches 13; Indels 0; Gaps 0;
QY 4 CAYGCNTGGCNCNTCCNTAYAAAGAYTAYGARGTNAAR 42
DB 82 CATGCTGGTCTCTCCCTAAGAGGATTAGGAAATAAAG 44
RESULT 86
AAD39736/c
ID AAD39736 standard; DNA; 548 BP.

```



Qy 4 CAYGCTGGCCNTCCNTAARGAYTAYGARGTNAAR 42  
ID AAC57760/c  
Db 81 CATGCTGGGTCTTCCCTAAGAGGATTAGGAATAAG 43

RESULT 88  
AAC57760/c  
ID AAC57760 standard; DNA; 1001 BP.  
XX AC AAC57760;  
XX AC AAC57760;  
XX DT 25-JAN-2001 (first entry)  
XX DE Arachidonic acid metabolism related genomic biallelic marker #394.

XX Human; biallelic marker; arachidonic acid metabolism; genotyping;  
KW detection; hybridisation; phenotype; haplotype; SNP; polymorphic base;  
KW single nucleotide polymorphism; hybridisation assay; sequencing assay;  
KW specific amplification assay; identification; ERM; 12-LO-RBM;  
KW eicosanoid-related biallelic marker; 12-LO-related biallelic marker; ds.  
XX Homo sapiens.  
XX OS  
XX PN W0200047771-A2.  
XX PD 17-AUG-2000.  
XX PF 11-FEB-2000; 2000WO-IB000184.  
XX PR 12-FEB-1999; 99US-0119917P.  
XX PR 23-MAR-1999; 99US-00275267.  
XX PR 07-MAY-1999; 99US-0133200P.  
XX PA (GEST ) GENSET.  
XX PI Blumenfeld M, Bougueleret L, Chumakov I;  
XX WPI; 2000-571881/53.  
XX DR Novel biallelic markers useful for detecting conditions and genotypes  
XX PT associated with arachidonic acid metabolism.  
XX PS Claim 13; Page 566; 802pp; English.

XX The present invention describes polynucleotides including biallelic  
CC markers derived from genes involved in arachidonic acid metabolism and  
CC from genomic regions flanking those genes. Methods from the present  
CC invention may be used to select individuals for clinical trials and  
CC predict responses to treatment with drugs. The polynucleotides may be  
CC used in hybridisation assays, sequencing assays and specific  
CC amplification assays for identifying an eicosanoid-related biallelic  
CC marker (ERBM) or 12-LO-related biallelic marker, and for amplifying a  
CC segment of nucleotides containing an ERBM. The polynucleotides are useful  
CC in diagnostic kits. The markers may be used to detect conditions and  
CC genotypes associated with arachidonic acid metabolism. AAC57367 to  
CC AAC58018 and AAB24019 and AAB24020 represent sequences used in the  
CC exemplification of the present invention. N.B. Polymorphic bases (single  
CC nucleotide polymorphisms also known as SNPs) in the polynucleotide  
CC sequences from the present invention have been given as their  
CC corresponding degenerate bases e.g. a polymorphic base of C or T has been  
XX given as Y  
XX Sequence 1001 BP; 316 A; 185 C; 196 G; 303 T; 0 U; 1 Other;  
XX Query Match 57.8%; Score 19.2; DB 3; Length 1001;  
XX Best Local Similarity 64.3%; Pred. No. 2.9e+02;  
XX Matches 18; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Qy 8 CMTGGCCNTCCNTAARGAYTAYCA 35  
Db 169 CATGGCCATCTTCTTATAGGACAATGA 142

RESULT 89  
ADO30325  
ID ADO30325 standard; CDNA; 1068 BP.  
XX AC ADO30325;  
XX AC ADO30325;  
XX DT 29-JUL-2004 (first entry)  
XX DE Mouse GPCR RAI3 polynucleotide, SEQ ID NO:1428.  
XX DE  
XX KW G protein-coupled receptor; GPCR; drug screening; diagnosis;  
KW transgenic mouse; neurological disorder; adrenal gland disorder;  
KW colon disorder; intestinal disorder; cardiovascular disorder;  
KW muscular disorder; blood disorder; immune disorder; bone disorder;  
KW joint disorder; metabolic disorder; nutritive disorder; cancer;  
KW kidney disorder; liver disorder; lung disorder; breast disorder;  
KW ovary disorder; uterus disorder; prostate disorder; testis disorder;  
KW skin disorder; stomach disorder; pancreas disorder; spleen disorder;  
KW thymus disorder; thyroid disorder; antiparkinsonian; antimanic;  
KW cytostatic; antiinflammatory; vasotropic; antidiarrhoeic; antidiabetic;  
KW CNS; central nervous system; respiratory; antidiarrhoeic; antidiabetic;  
KW virucide; hepatotropic; antibacterial; antianemic; antiseborrhoeic;  
KW dermatological; antitumor; antithyroid; antiallergic; anorectic;  
KW immunosuppressive; nephrotropic; gene therapy; GPCR modulator; mouse;  
KW murine; gene; ss.  
XX OS  
XX MS musculus.  
XX PN W02004040000-A2.  
XX PD 13-MAY-2004.  
XX PF 09-SEP-2003; 2003WO-US028226.  
XX PR 09-SEP-2002; 2002US-0409303P.  
XX PR 09-APR-2003; 2003US-0461329P.  
XX PA (PRIM-) PRIMAL INC.  
XX PI Gaitanaris GA, Bergmann JE, Gragerov A, Hohmann J, Li F;  
XX Madisen L, McIlwain KL, Pavlova MN, Vassiliadis D, Zeng H;  
XX WPI; 2004-390329/36.  
XX P-PSDB; ADO29633.  
XX Novel mammalian G protein coupled receptors, useful for identifying  
XX compounds that modulates diagnosing and treating disease condition  
XX associated with GPCR dysfunction e.g. autoimmune diseases, angina  
XX pectoris, Parkinson's disease.  
XX Claim 151; SEQ ID NO 1428; 542pp; English.  
XX The invention relates to human and mouse G protein-coupled receptors  
XX (GPCRs) and nucleic acids encoding them. The invention also relates to  
XX sequences at least 90% identical to the GPCR proteins and nucleic acids  
XX of the invention; methods of treating, preventing or diagnosing diseases  
XX associated with GPCRs of the invention; methods of screening for  
XX compounds useful in the treatment of GPCR-related diseases; a transgenic  
XX mouse comprising a GPCR gene of the invention; a mouse comprising a  
XX mutation in a GPCR transgene or in an endogenous GPCR gene; cells derived  
XX from the transgenic mice; kits comprising several mice, each of which has  
XX a mutation in a different GPCR gene of the invention; and kits comprising  
XX probes which hybridise to GPCR polynucleotides of the invention. The  
XX invention further discloses variants of the GPCR polypeptides and vectors  
XX comprising a GPCR nucleic acid. The GPCR nucleic acids and proteins may  
XX be used in the diagnosis, treatment or prevention of a wide variety of  
XX diseases including neurological disorders (e.g., Alzheimer's disease,  
XX depression, diabetic neuropathy, Parkinson's disease or schizophrenia);  
XX disorders of the adrenal gland; disorders of the colon or intestine  
XX (e.g., Crohn's disease, diarrhoea, food poisoning or irritable bowel  
XX syndrome); cardiovascular disorders (e.g., angina, cardiac arrhythmia or  
XX myocardial infarction); muscular disorders; blood disorders (e.g.,  
XX anaemia or leukaemia); immune disorders (e.g., autoimmune disorders or





DE Chlamydia trachomatis antigen polynucleotide SEQ ID NO:6.  
XX Chlamydia infection; Chlamydia; antibiotic; antiinflammatory;  
KW antiinfertility; cardiant; antiarteriosclerotic; ophthalmological;  
KW vaccine; gene therapy; immune response; pelvic inflammatory disease;  
KW tubal obstruction; infertility; male infertility; ocular infection;  
KW blindness; acute respiratory tract infection; atherosclerosis;  
KW coronary heart disease; gene; ds.  
XX Chlamydia trachomatis.  
XX Chlamydia trachomatis.  
XX WO2003041560-A2.  
XX 22-MAY-2003.  
XX 05-NOV-2002; 2002WO-US035624.  
XX 06-NOV-2001; 2001US-00012256.  
XX 05-DEC-2001; 2001US-00007693.  
XX 15-JUL-2002; 2002US-0019720.  
XX (CORI-) CORIXA CORP.  
XX Bhatia A, Guderian J, Skeiky YAW, Maisonneuve JL, Barth B;  
PI Probst P;  
XX WPI; 2003-441771/41.  
XX New DNA and proteins comprising a portion of a Chlamydia antigen, useful  
PT for diagnosing or treating Chlamydial infections, particularly as  
PT vaccines for treating or preventing Chlamydial infections, e.g. pelvic  
PT inflammatory disease.  
XX Example 1; SEQ ID NO 6; 275pp; English.  
XX The present invention describes compounds and methods for diagnosing and  
CC treating Chlamydia infection. Chlamydia polynucleotide and protein  
CC sequences have antibiotic, antiinflammatory, antiinfertility, cardiant,  
CC antiarteriosclerotic and ophthalmological activities, and can be used in  
CC vaccines, and in gene therapy. The Chlamydia polynucleotides, proteins,  
CC compositions or methods from the present invention can be used for the  
CC serodiagnosis or treatment of Chlamydial infections, particularly in  
CC humans. The polynucleotides, proteins or compositions are particularly  
CC useful for stimulating an immune response in a patient, or for  
CC stimulating and/or expanding T cells specific for a Chlamydia protein.  
CC Specifically, the polynucleotides, proteins or compositions are useful as  
CC vaccines for treating or preventing Chlamydial infections including  
CC pelvic inflammatory disease (which results in tubal obstruction and  
CC infertility in women), male infertility, ocular infection (which may  
CC cause blindness), acute respiratory tract infections, atherosclerosis, or  
CC coronary heart disease. The present sequence is used in the  
CC exemplification of the present invention.  
XX Sequence 1828 BP; 571 A; 449 C; 317 G; 491 T; 0 U; 0 Other;  
Query Match 57.8%; Score 19.2; DB 10; Length 1828;  
Best Local Similarity 56.8%; Pred. No. 3.4e+02;  
Matches 21; Conservative 4; Mismatches 12; Indels 0; Gaps 0;  
QY 6 YGCGTGGCCNTCCNTAYAAAGAYTAYGARGTNAAR 42  
DB 1456 TGCTGGCGCTCTCCGTAGAACACTTTATGTATA 1492  
RESULT 93  
ABL12999  
ID ABL12999 standard; cDNA; 2090 BP.  
XX ABL12999;  
XX 26-MAR-2002 (first entry)  
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 33479.  
DE

XX Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX Drosophila melanogaster.  
XX WO200171042-A2.  
XX 27-SEP-2001.  
XX 23-MAR-2001; 2001WO-US009231.  
XX 23-MAR-2000; 2000US-0191637P.  
XX 11-JUL-2000; 2000US-00614150.  
XX (PEKE ) PE CORP NY.  
XX Venter JC, Adams M, Li PWD, Myers EW;  
PI WPI; 2001-656860/75.  
XX P-PSDB; ABB68896.  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.  
XX Claim 1; SEQ ID NO 33479; 21pp + Sequence Listing; English.  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX Sequence 2090 BP; 489 A; 579 C; 599 G; 423 T; 0 U; 0 Other;  
Query Match 57.8%; Score 19.2; DB 4; Length 2090;  
Best Local Similarity 50.0%; Pred. No. 3.5e+02;  
Matches 21; Conservative 6; Mismatches 15; Indels 0; Gaps 0;  
QY 1 GCNCAYGCGTGGCCNTCCNTAYAAAGAYTAYGARGTNAAR 42  
DB 1257 GCGCAGCTCTGGATGGATCAGTACAAGGATTACTATCTCAAG 1298  
RESULT 94  
ADL12908/C  
ID ADL12908 standard; cDNA; 2496 BP.  
XX ADL12908;  
XX 06-MAY-2004 (first entry)  
XX Human steroid-induced C3A liver cell cDNA #637.  
XX ss; gene; Hepatotropic; Gene therapy; Wilson disease; liver disorder;  
KW steroid therapy; cirrhosis; hepatitis; human; C3A liver cell.  
XX Homo sapiens.  
XX US6673549-B1.  
XX 06-JAN-2004.  
XX 12-OCT-2001; 2001US-00976594.  
XX 12-OCT-2000; 2000US-0240409P.  
XX

```
PA (INCY-) INCYTE CORP.
XX Furness LM, Buchbinder JL;
XX WPI; 2004-068610/07.
XX
XX Combination useful for preparing a composition for treating liver
PT disorders associated with steroid therapy, e.g., cirrhosis or hepatitis,
PT comprises cDNAs that are differentially expressed in response to steroid
PT treatment.
XX
XX Claim 1; SEQ ID NO 637; 141pp; English.
XX
XX The invention relates to a combination comprising cDNAs that are
XX differentially expressed in response to steroid treatment. Also included
XX are the following: a high throughput method for using a cDNA to detect
XX differential expression of nucleic acids in a sample; and a high
XX throughput method of screening molecules or compounds to identify a
XX ligand that specifically binds a cDNA. The sample is from a subject with
XX Wilson disease and comparison of a standard defines a stage of that
XX disease. The high throughput method of screening molecules or compounds
XX to identify a ligand that specifically binds a cDNA comprises: combining
XX the combination with molecules or compounds under conditions to allow
XX specific binding; and detecting specific binding between each cDNA and at
XX least one molecule or compound. The molecules or compounds are regulatory
XX proteins. The combination is useful for preparing a composition for
XX treating liver disorders associated with steroid therapy, e.g., cirrhosis
XX or hepatitis. The present sequence represents a human cDNA which is
XX differentially expressed in steroid-induced C3A liver cells. Note: The
XX sequence data for this patent did not form part of the printed
XX specification but was obtained in electronic format directly from USPTO
XX at seqdata.uspto.gov/sequence.html.
XX
XX SQ Sequence 2496 BP; 867 A; 344 C; 457 G; 828 T; 0 U; 0 Other;

Query Match 57.8%; Score 19.2; DB 12; Length 2496;
Best Local Similarity 56.8%; Pred. No. 3.7e+02;
Matches 21; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

QY 5 AYCNGTGGCNCNTCCNTAYAAAGATYATYAGCTNAA 41
   |||||
Db 1130 ACGCTGGACTTCTCAAATATGACATGAGTCAA 1094
   |||||

RESULT 95
AAS34914/c
ID AAS34914 standard; cDNA; 2648 BP.
XX
XX AAS34914;
AC
XX
XX 04-DEC-2001 (first entry)
DT
XX
XX cDNA encoding novel human neoplastic disease associated polypeptide #148.
DE
XX
XX Human; neoplastic disease associated polypeptide; cancer; gene therapy;
KW hyperproliferative disorder; neural disorder; immune system disorder;
KW muscular disorder; reproductive disorder; gastrointestinal disorder;
KW pulmonary disorder; cardiovascular disorder; renal disorder;
KW neuroprotective; cytostatic; anti inflammatory; vasotropic; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200155163-A1.
PN
XX
XX 02-AUG-2001.
PD
XX
XX 17-JAN-2001; 2001WO-US0001358.
PF
XX
XX 31-JAN-2000; 2000US-0179065P.
PR
XX 04-FEB-2000; 2000US-0180628P.
PR
XX 24-FEB-2000; 2000US-0184664P.
PR
XX 02-MAR-2000; 2000US-0186350P.
PR
XX 16-MAR-2000; 2000US-0189874P.
PR
17-MAR-2000; 2000US-0190076P.
PR
18-APR-2000; 2000US-0198123P.
PR
19-MAY-2000; 2000US-0205515P.
PR
07-JUN-2000; 2000US-0209467P.
PR
28-JUN-2000; 2000US-0214886P.
PR
30-JUN-2000; 2000US-0215135P.
PR
07-JUL-2000; 2000US-0216647P.
PR
07-JUL-2000; 2000US-0216880P.
PR
11-JUL-2000; 2000US-0217487P.
PR
11-JUL-2000; 2000US-0217496P.
PR
14-JUL-2000; 2000US-0218290P.
PR
26-JUL-2000; 2000US-0220963P.
PR
26-JUL-2000; 2000US-0220964P.
PR
14-AUG-2000; 2000US-0224518P.
PR
14-AUG-2000; 2000US-0224519P.
PR
14-AUG-2000; 2000US-0225213P.
PR
14-AUG-2000; 2000US-0225214P.
PR
14-AUG-2000; 2000US-0225266P.
PR
14-AUG-2000; 2000US-0225267P.
PR
14-AUG-2000; 2000US-0225268P.
PR
14-AUG-2000; 2000US-0225270P.
PR
14-AUG-2000; 2000US-0225447P.
PR
14-AUG-2000; 2000US-0225757P.
PR
14-AUG-2000; 2000US-0225758P.
PR
14-AUG-2000; 2000US-0225759P.
PR
18-AUG-2000; 2000US-0226279P.
PR
22-AUG-2000; 2000US-0226681P.
PR
22-AUG-2000; 2000US-0226868P.
PR
22-AUG-2000; 2000US-0227182P.
PR
23-AUG-2000; 2000US-0227009P.
PR
30-AUG-2000; 2000US-0228924P.
PR
01-SEP-2000; 2000US-0229287P.
PR
01-SEP-2000; 2000US-0229343P.
PR
01-SEP-2000; 2000US-0229344P.
PR
01-SEP-2000; 2000US-0229345P.
PR
05-SEP-2000; 2000US-0229509P.
PR
05-SEP-2000; 2000US-0229513P.
PR
06-SEP-2000; 2000US-0230437P.
PR
06-SEP-2000; 2000US-0230438P.
PR
08-SEP-2000; 2000US-0231242P.
PR
08-SEP-2000; 2000US-0231243P.
PR
08-SEP-2000; 2000US-0231244P.
PR
08-SEP-2000; 2000US-0231413P.
PR
08-SEP-2000; 2000US-0231414P.
PR
08-SEP-2000; 2000US-0232080P.
PR
08-SEP-2000; 2000US-0232081P.
PR
12-SEP-2000; 2000US-0231968P.
PR
14-SEP-2000; 2000US-0232397P.
PR
14-SEP-2000; 2000US-0232398P.
PR
14-SEP-2000; 2000US-0232399P.
PR
14-SEP-2000; 2000US-0232400P.
PR
14-SEP-2000; 2000US-0232401P.
PR
14-SEP-2000; 2000US-0233063P.
PR
14-SEP-2000; 2000US-0233064P.
PR
14-SEP-2000; 2000US-0233065P.
PR
21-SEP-2000; 2000US-0234223P.
PR
21-SEP-2000; 2000US-0234274P.
PR
25-SEP-2000; 2000US-0234997P.
PR
25-SEP-2000; 2000US-0234998P.
PR
26-SEP-2000; 2000US-0235484P.
PR
27-SEP-2000; 2000US-0235834P.
PR
27-SEP-2000; 2000US-0235836P.
PR
29-SEP-2000; 2000US-0236327P.
PR
29-SEP-2000; 2000US-0236327P.
PR
29-SEP-2000; 2000US-0236367P.
PR
29-SEP-2000; 2000US-0236369P.
PR
29-SEP-2000; 2000US-0236370P.
PR
02-OCT-2000; 2000US-0236802P.
PR
02-OCT-2000; 2000US-0237037P.
PR
02-OCT-2000; 2000US-0237038P.
PR
02-OCT-2000; 2000US-0237039P.
PR
02-OCT-2000; 2000US-0237040P.
PR
13-OCT-2000; 2000US-0239935P.
```





CC condition or a susceptibility to a pathological condition in a subject,  
 CC identifying a binding partner to the polypeptide, identifying an activity  
 CC in a biological assay, and the gene corresponding to the cDNA sequence.  
 CC The polypeptides, polynucleotides and antibodies are useful for  
 CC detecting, preventing, diagnosing, prognosticating, treating or  
 CC ameliorating medical conditions such as hyperproliferative diseases or  
 CC cancer, autoimmune disorders (e.g. diabetes, rheumatoid arthritis,  
 CC systemic lupus erythematosus, multiple sclerosis, autoimmune thyroiditis  
 CC or haemolytic anaemia), haematopoietic or haematologic disorders (e.g.  
 CC anaemia of thrombocytopaenia), allergic reactions including asthma or  
 CC eczema, inflammatory disorders (e.g. ischaemia-reperfusion injury,

Query Match 57.8%; Score 19.2; DB 10; Length 2648;  
 Best Local Similarity 56.8%; Pred. No. 3.8e+02;  
 Matches 21; Conservative 4; Mismatches 12; Indels 0; Gaps 0;  
 Qy 5 AYGNTGGCNCNTCCNTAAYARGAYTAYGARGTNA 41  
 Db 1245 ACGCCTGGACTTCTCAAAATTTATGAGCATGAAGTCAA 1209

RESULT 97  
 ACA30557/c

ID ACA30557 standard; DNA; 3096 BP.

AC ACA30557;

DT 19-JUN-2003 (first entry)

XX Prokaryotic essential gene #12214.

XX Antisense; ds; prokaryotic essential gene; cell proliferation;  
 XX drug design; gene.

XX Campylobacter jejuni.

XX WO200277183-A2.

XX 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00815242.

XX 06-SEP-2001; 2001US-00948993.

XX 23-OCT-2001; 2001US-0342923P.

XX 08-FEB-2002; 2002US-00072851.

XX 06-MAR-2002; 2002US-0362659P.

XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen XL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GU, Yamamoto R, Forsyth RA, Xu HH;

XX WPI; 2003-029926/02.

XX P-PSDB; ABU26687.

XX New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.

XX Claim 14; SEQ ID NO 18427; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of

CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
 CC prokaryotic essential genes. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 3096 BP; 1189 A; 420 C; 642 G; 845 T; 0 U; 0 Other;

Query Match 57.8%; Score 19.2; DB 8; Length 3096;

Best Local Similarity 52.5%; Pred. No. 4e+02;

Matches 21; Conservative 5; Mismatches 14; Indels 0; Gaps 0;

Qy 1 GCNCAYGCNTGCCNTCCNTAAYARGAYTAYGARGTNA 40

Db 1978 GCCCTTCATAGGCTTATCTATATATAAGAGTATGAGCGA 1939

RESULT 98

ADV35091

ID ADV35091 standard; cDNA; 3397 BP.

AC ADV35091;

XX 10-FEB-2005 (first entry)

XX Rat cDNA differentially expressed in the presence of VIP SeqID167.

XX rat; VIP; vasoactive intestinal polypeptide; ss;

XX multi-parameter high throughput screening; MPHTS; disease signature;

XX neuropsychiatric; neurodegenerative; schizophrenia;

XX bipolar affective disorder; BAD; autism; Parkinson's;

XX Alzheimer's disease; neuroleptic; nootropic; antimanic; antidepressant.

XX Rattus norvegicus.

XX US2003096264-A1.

XX 22-MAY-2003.

XX 18-JUN-2002; 2002US-00175523.

XX 18-JUN-2001; 2001US-0299151P.

XX 07-SEP-2001; 2001US-0317828P.

XX 25-SEP-2001; 2001US-0325150P.

XX 14-NOV-2001; 2001US-0333047P.

XX 18-JAN-2002; 2002US-0349936P.

XX 04-MAR-2002; 2002US-0361834P.

XX (PSYC-) PSYCHIATRIC GENOMICS INC.

XX Altar CA, Brockman JA, Evans D, Hook D, Klimczak LJ, Laeng P;  
 PI Palfreyman M, Rajan P;  
 XX WPI; 2004-118903/12.  
 XX Identifying a compound that can treat disease or disorders, such as, a  
 PT neuropsychiatric disorder e.g., schizophrenia, or autism, comprises  
 PT determining the expression of one or more efficacy genes in a cell

```
PT contacted with the test compound.
XX Claim 10; SEQ ID NO 167; 39pp; English.
XX This invention relates to a novel screening method identified as a multi-
CC parameter high throughput screening (MPHTS) assay. Specifically, it
CC refers to an assay that utilizes the disease signature of a plurality of
CC specific genes associated with a particular disease, and identifies
CC differential expression between those cells taken from individuals
CC affected by that disease and those that are not affected. The present
CC invention then describes the screening of candidate pharmaceutical
CC compounds to identify those that have a potential therapeutic benefit for
CC the treatment of neuropsychiatric and neurodegenerative disorders
CC including schizophrenia, bipolar affective disorder (BAD) and autism, as
CC well as Parkinson's and Alzheimer's disease. Accordingly, the compounds
CC of this invention exhibit various activities including neuroleptic,
CC nootropic, antimanic and antidepressant. Furthermore, the screening
CC method used in MPHTS will be automated, such that a large number of test
CC compounds may be rapidly screened with a minimal amount of labour and
CC effort. This polynucleotide is a cDNA sequence of a gene that is
CC differentially expressed in rats in the presence of the therapeutic
CC compound VIP (vasoactive intestinal polypeptide) given in an
CC exemplification of the invention.
XX
XX Sequence 3397 BP; 711 A; 1028 C; 959 G; 699 T; 0 U; 0 Other;
SQ
Query Match 57.8%; Score 19.2; DB 13; Length 3397;
Best Local Similarity 52.5%; Pred. No. 4.1e+02;
Matches 21; Conservative 5; Mismatches 14; Indels 0; Gaps 0;
QY 2 CNCAYGCNTGGCCNTCCNTAYAAAGAYTAYGARGCTNAA 41
DB 2452 CGGATGAAGGGTCTCCCTCTACCACGCTCTATGAGGTGAA 2491
RESULT 99
ADZ60766 ID ADZ60766 standard; DNA; 3397 BP.
XX AC ADZ60766;
XX 30-JUN-2005 (first entry)
XX Rat Ptprn gene, SEQ ID 490.
XX Drug screening; Peroxisome Proliferator-Activated Receptor gamma;
KW PPAR gamma; gene; ds.
XX Rattus norvegicus.
XX US2005084872-A1.
XX 21-APR-2005.
XX 23-JAN-2004; 2004US-00764420.
XX 24-JAN-2003; 2003US-0442797P.
XX 30-MAY-2003; 2003US-0474413P.
XX (LUMP/) LUM P Y.
XX (TANY/) TAN Y.
XX (DAIH/) DAI H.
XX (MUIS/) MUISE E S.
XX (BERG/) BERGER J P.
XX (THOM/) THOMPSON J R.
XX Lum PY, Tan Y, Dai H, Muise ES, Berger JP, Thompson JR;
XX WPI; 2005-313963/32.
XX GENBANK; U40652.
XX Determining whether agent has defined biological activity, by making
XX comparison e.g. comparing efficacy value, toxicity value and classifier
PT
```

PR 15-MAR-2002; 2002US-0364134P.  
PR 08-APR-2002; 2002US-0370144P.  
PR 08-APR-2002; 2002US-0370206P.  
PR 08-APR-2002; 2002US-0370247P.  
PR 17-APR-2002; 2002US-0372794P.  
PR 21-APR-2002; 2002US-0371679P.  
XX  
PA (GENE-) GENE LOGIC INC.  
XX  
XX Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Elashoff M;  
XX WPI; 2003-148464/14.  
DR  
XX  
XX Predicting at least one toxic effect of a compound, useful for toxicity  
PT modeling, comprises preparing a gene expression profile of a tissue or  
PT cell sample exposed to the compound, and comparing the gene expression  
PT profile to a database.  
XX  
XX Example 4; Page; 446pp; English.  
PS  
XX The invention relates to a novel method of predicting at least one toxic  
CC effect of a compound. The method comprises a gene expression profile of a  
CC tissue or cell sample exposed to the compound, and comparing the gene  
CC expression profile to a database comprising at least part of the data or  
CC information given in the specification. The methods are useful for  
CC predicting at least one toxic effect of a compound, predicting the  
CC progression of a toxic effect of a compound, predicting the renal  
CC toxicity of a compound, or identifying toxicity markers in tissues or  
CC cells exposed to known renal toxin. The genes are useful as toxicity  
CC markers in drug screening and toxicity assays, in monitoring disease or  
CC physiological states, or disease progression. This polynucleotide  
CC represents a rat DNA sequence relating to the toxic effect database  
CC described in the specification. NOTE: The sequence data for this patent  
CC did not form part of the printed specification, but was obtained in  
CC electronic format directly from the World Intellectual Property  
CC Organization  
XX  
SQ Sequence 3477 BP; 720 A; 1050 C; 998 G; 709 T; 0 U; 0 Other;  
  
Query Match 57.8%; Score 19.2; DB 10; Length 3477;  
Best Local Similarity 52.5%; Pred. No. 4.1e+02;  
Matches 21; Conservative 5; Mismatches 14; Indels 0; Gaps 0;  
  
OY 2 CNCAYCNCNTGCCNCCNTAARGAYTAYGARGTNA 41  
DB 2510 CGGATGAAGGGTCTCTCCCTCTACCACGCTCTATGAGGTGAA 2549

Search completed: June 10, 2006, 15:33:27  
Job time : 348.45 secs

**THIS PAGE IS BLANK**



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2006, 15:19:28 ; Search time 5182.8 seconds  
(without alignments)  
453.155 Million cell updates/sec

Title: US-10-600-816-33

Perfect score: 33.2

Sequence: 1 gcncaaycgtggcncntccc.....ayaargaytaygargtnaar 42

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 48236798 seqs, 27959665780 residues

Total number of hits satisfying chosen parameters: 96473596

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

Database :

EST:\*

- 1: gb\_est1:\*
- 2: gb\_est3:\*
- 3: gb\_est4:\*
- 4: gb\_est5:\*
- 5: gb\_est6:\*
- 6: gb\_est7:\*
- 7: gb\_est8:\*
- 8: gb\_est9:\*
- 9: gb\_est10:\*
- 10: gb\_est11:\*
- 11: gb\_gss1:\*
- 12: gb\_gss2:\*
- 13: gb\_gss3:\*
- 14: gb\_gss4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	30	90.4	158	2	BI063311 IL3-UT011
C 2	30	90.4	307	7	BE818833
C 3	30	90.4	341	7	BE818817
C 4	30	90.4	384	7	BE818835
C 5	30	90.4	396	7	BE818828
C 6	30	90.4	404	7	BE818840
C 7	30	90.4	424	7	BE818805
C 8	30	90.4	449	7	BE818807
C 9	30	90.4	453	5	CD672930
C 10	30	90.4	458	7	BE818871
C 11	30	90.4	466	7	BE818859
C 12	30	90.4	467	7	BE818852
C 13	30	90.4	571	3	BP253378
C 14	30	90.4	577	4	CB152636
C 15	30	90.4	580	3	BP252334
C 16	30	90.4	580	3	BP267627
C 17	30	90.4	580	3	BP268151
C 18	30	90.4	582	3	BP256144
C 19	30	90.4	582	3	BP363758





C 385 19.6 59.0 728 13 CW354811  
C 386 19.6 59.0 730 3 BU349060  
C 387 19.6 59.0 733 2 BU299093  
C 388 19.6 59.0 735 10 DV953868  
C 389 19.6 59.0 739 3 BU293342  
C 390 19.6 59.0 740 1 AJ614305  
C 391 19.6 59.0 743 7 BE415805  
C 392 19.6 59.0 747 13 CL911455  
C 393 19.6 59.0 749 11 BZ003523  
C 394 19.6 59.0 750 12 CG865087  
C 395 19.6 59.0 751 14 AG393824  
C 396 19.6 59.0 752 11 BH387047  
C 397 19.6 59.0 757 8 COS18016  
C 398 19.6 59.0 759 12 CG283327  
C 399 19.6 59.0 766 10 DR792922  
C 400 19.6 59.0 768 13 CZ808411  
C 401 19.6 59.0 772 13 CZ271612  
C 402 19.6 59.0 774 10 DR799211  
C 403 19.6 59.0 778 4 CA196818  
C 404 19.6 59.0 779 7 BE908400  
C 405 19.6 59.0 783 10 DR965259  
C 406 19.6 59.0 784 2 BI829579  
C 407 19.6 59.0 796 13 CW354812  
C 408 19.6 59.0 809 3 BU227668  
C 409 19.6 59.0 812 4 CB235984  
C 410 19.6 59.0 815 7 BE194336  
C 411 19.6 59.0 816 3 BU119565  
C 412 19.6 59.0 825 10 DV173400  
C 413 19.6 59.0 837 3 BU239087  
C 414 19.6 59.0 840 11 BH403653  
C 415 19.6 59.0 843 8 CN148798  
C 416 19.6 59.0 852 10 DR806206  
C 417 19.6 59.0 872 13 CZ861141  
C 418 19.6 59.0 874 13 CZ697321  
C 419 19.6 59.0 882 8 CO815184  
C 420 19.6 59.0 912 4 BX454774  
C 421 19.6 59.0 918 12 CG939142  
C 422 19.6 59.0 935 14 DU736675  
C 423 19.6 59.0 942 12 CC416069  
C 424 19.6 59.0 948 10 DW612643  
C 425 19.6 59.0 966 14 CNU015SL  
C 426 19.6 59.0 992 13 DU092813  
C 427 19.6 59.0 1023 14 CNU049WR  
C 428 19.6 59.0 1095 14 CNU012K0  
C 429 19.6 59.0 1155 9 DN733421  
C 430 19.6 59.0 1321 8 CV860693  
C 431 19.6 59.0 1464 2 BI458506  
C 432 19.6 59.0 2951 6 AK039029  
C 433 19.6 59.0 2952 6 AK138499  
C 434 19.6 59.0 4091 12 CG865108  
C 435 19.4 58.4 156 7 AW865253  
C 436 19.4 58.4 218 8 CX028512  
C 437 19.4 58.4 263 1 AO1023526  
C 438 19.4 58.4 288 1 AA815461  
C 439 19.4 58.4 286 1 AA459692  
C 440 19.4 58.4 306 1 AA910794  
C 441 19.4 58.4 342 4 CB158087  
C 442 19.4 58.4 343 4 CA367145  
C 443 19.4 58.4 359 9 CX624624  
C 444 19.4 58.4 363 14 DU998011  
C 445 19.4 58.4 364 10 DV535159  
C 446 19.4 58.4 367 1 AO1005427  
C 447 19.4 58.4 367 3 BP430822  
C 448 19.4 58.4 367 12 CE541719  
C 449 19.4 58.4 368 11 AZ971882  
C 450 19.4 58.4 376 3 BP431446  
C 451 19.4 58.4 402 13 DU161690  
C 452 19.4 58.4 411 14 DU814274  
C 453 19.4 58.4 416 1 AA609420  
C 454 19.4 58.4 434 4 CA382669  
C 455 19.4 58.4 435 4 CA329264  
C 456 19.4 58.4 439 8 CO515796  
C 457 19.4 58.4 459 7 BF641113

CW354811 f5bb001f0  
BU349060 604171370  
BU299093 BU299093  
DV953868 SB03027B1  
BU293342 603323255  
AJ614305 AJ614305  
BE415805 MUG001.CO  
CL911455 OA\_Aba001  
BZ003523 oeh91f09.  
CG865087 116D149P6  
AG393824 Mus muscu  
BH387047 AG-ND-179  
COS18016 3530\_1\_11  
CG283327 OG5DM367C  
DR792922 ZM\_BFB001  
CZ808411 OC\_Ba018  
CZ271612 ZMMEF0051  
DR799211 ZM\_BFB002  
CA196818 SCBFAD109  
BE908400 601503023  
DR965259 ZM\_BFB008  
BI829579 603079315  
CZ271612 f5bb001f0  
BU227668 603797338  
CB235984 AGENCOURT  
BE194336 HVSMEH008  
BU119565 603142825  
DV173400 ZM\_BFB017  
BU239087 603323967  
BH403653 AG-ND-146  
CN148798 WOUND1.56  
DR806206 ZM\_BFB003  
CZ861141 OC\_Ba025  
CZ697321 OC\_Ba001  
CO815184 AGENCOURT  
BX454774 BX454774  
CG939142 MBELUS91PR  
DU736675 APK13883.  
CC416069 PUHQG657D  
DW612643 CLJ270-CO  
AL105855 Drosophil  
DU92813 104276 To  
AL281124 Tetraodon  
AL101658 Drosophil  
DN733421 CNB70-A01  
CV860693 gonad\_EST  
BI458506 603198896  
AK039029 Mus muscu  
AK138499 Mus muscu  
CG865108 phage3741  
AW865253 PM4-SN002  
CX028512 1340277 N  
AO1023526 ov79G04.s  
AA815461 a165e02.s  
AA459692 ZXA9e07.s  
AA910794 O148C08.s  
CB158087 K-EST0217  
CA367145 642894 NC  
CX624624 A24\_A12.S  
DU998011 OG\_BBa013  
DV535159 ZM\_BFB022  
AO1005427 ou08C03.x  
BP430822 BP430822  
CE541719 t1gr-gss-  
AZ971882 ZM0245J02  
BP431446 BP431446  
DU161690 OG\_Aba004  
DU814274 OG\_Aba010  
AA609420 zu72C06.s  
CA382669 662525 NC  
CA329264 hab23e03.  
CO515796 e13dSG54B  
BF641113 NF032D031

BE502143 hv13B09.x  
BX675014 BX675014  
AW044486 wx22e06.x  
WS514161 hd24e04.x  
DX235462 OR\_Aba011  
BM852382 K-EST0133  
BE501526 hw33c11.x  
AI038249 oy85e05.x  
AA427486 zw26e04.x  
AJ709346 AJ709346  
DX196616 OR\_Aba006  
DX306704 OR\_Aba021  
AI118590 us94e06.x  
CA374622 649086 NC  
AI640610 wal6b05.x  
CK875130 SGP136355  
DU148599 OG\_Aba001  
DX202775 OR\_Aba007  
CF793715 889149 MA  
AI082814 ox78a07.x  
AI672545 wp32e05.x  
CD672812 fg16e06.y  
AI623831 t623g12.x  
AA813357 a181d05.s  
AI190291 q357e04.x  
CK866657 SGP148525  
DB237826 DB237826  
BE339690 601310454  
DX233198 OR\_Aba011  
DX305420 OR\_Aba021  
DV951545 SB03020B1  
BZ249649 CH230-447  
DU816826 OG\_Aba009  
DX227932 OR\_Aba010  
AI913001 wa09h07.x  
DX235689 OR\_Aba011  
AI652099 wp27h09.x  
DT672790 a13dtU44H  
AZ116714 RPl-23-2  
CL920572 OA\_Aba002  
DX263465 OR\_Aba015  
AJ707259 AJ707259  
BJ668716 BJ668716  
DT671238 a13dtU28A  
CG812627 FSAAK74TR  
DX183948 OR\_Aba004  
DX202917 OR\_Aba007  
DA419705 DA419705  
BJ677748 BJ677748  
DX209586 OR\_Aba008  
CA382276 662043 NC  
DX219250 OR\_Aba009  
DX191898 OR\_Aba005  
DX217929 OR\_Aba009  
DX250186 OR\_Aba013  
DU812605 OC\_Aba008  
AI907844 RC-BT155-  
AI872397 wh58d07.x  
BQ009901 UI-H-ED1-  
BU632322 UI-H-FE1-  
DX23889 OR\_Aba010  
DX217111 OR\_Aba016  
CN860225 000816AAF  
CN261334 170006001  
DX246387 OR\_Aba013  
DW249226 Cm\_mxl\_20  
CZ570265 OR\_Ba000  
BQ015866 UI-H-D10-  
DX174659 OR\_Aba003  
CW255829 104\_720\_1  
DX244630 OR\_Aba012  
AI907902 RC-BT155-  
DX273859 OR\_Aba017

458 19.4 58.4  
C 459 19.4 58.4  
460 19.4 58.4  
461 19.4 58.4  
462 19.4 58.4  
463 19.4 58.4  
464 19.4 58.4  
465 19.4 58.4  
466 19.4 58.4  
467 19.4 58.4  
468 19.4 58.4  
469 19.4 58.4  
470 19.4 58.4  
471 19.4 58.4  
472 19.4 58.4  
473 19.4 58.4  
474 19.4 58.4  
475 19.4 58.4  
476 19.4 58.4  
477 19.4 58.4  
478 19.4 58.4  
483 19.4 58.4  
485 5 CK875130  
485 13 DU148599  
487 14 DX202775  
493 5 CF793715  
496 1 AI082814  
497 1 AI672545  
508 5 CD672812  
513 1 AI623831  
518 1 AA813357  
518 1 AI190291  
518 5 CK866657  
518 9 DB237826  
520 7 BE339690  
520 14 DX233198  
521 14 DX305420  
528 10 DV951545  
528 11 BZ249649  
528 14 DU816826  
530 14 DX227932  
531 1 AI913001  
535 14 DX235689  
538 1 AI652099  
541 10 DT672790  
541 11 AZ116714  
550 13 CL920572  
550 14 DX263465  
551 14 AJ707259  
552 2 BJ668716  
552 10 DT671238  
555 12 CG812627  
557 14 DX183948  
558 14 DX202917  
560 9 DA419705  
563 2 BJ677748  
563 14 DX209586  
580 4 CA382276  
582 14 DX219250  
584 14 DX191898  
585 14 DX217929  
585 14 DX250186  
587 14 DU812605  
589 1 AI907844  
590 1 AI872397  
593 3 BQ009901  
594 3 BU632322  
595 14 DX23889  
595 14 DX217111  
596 8 CN860225  
601 8 CN261334  
601 14 DX246387  
604 10 DW249226  
606 13 CZ570265  
608 3 BQ015866  
608 14 DX174659  
609 13 CW255829  
612 14 DX244630  
614 1 AI907902  
616 14 DX273859







```

c 969      19 57.2      550 9 DA742945
c 970      19 57.2      550 9 DB219745
c 971      19 57.2      550 9 DB285983
c 972      19 57.2      551 9 DA800048
c 973      19 57.2      551 12 CC559317
c 974      19 57.2      552 9 DA652161
c 975      19 57.2      553 9 DB077106
c 976      19 57.2      553 9 DB237760
c 977      19 57.2      554 9 DA860830
c 978      19 57.2      554 11 A2718342
c 979      19 57.2      556 9 DA222240
c 980      19 57.2      556 9 DA406387
c 981      19 57.2      557 9 DA242298
c 982      19 57.2      558 9 DA117988
c 983      19 57.2      558 9 DB165266
c 984      19 57.2      558 10 DT026877
c 985      19 57.2      559 12 CE447270
c 986      19 57.2      560 9 DA757655
c 987      19 57.2      561 9 DA716058
c 988      19 57.2      562 8 CV913151
c 989      19 57.2      562 9 DA726057
c 990      19 57.2      563 9 DA587461
c 991      19 57.2      563 9 DA721371
c 992      19 57.2      564 9 DA720857
c 993      19 57.2      564 9 DA724206
c 994      19 57.2      565 3 BU473020
c 995      19 57.2      565 9 DA725571
c 996      19 57.2      565 9 DB123941
c 997      19 57.2      565 13 CZ035288
c 998      19 57.2      565 14 DX138293
c 999      19 57.2      566 9 DA460877
c1000     19 57.2      566 9 DA525595

ALIGNMENTS

RESULT 1
BI063311      158 bp      mRNA      linear      EST 15-JUN-2001
LOCUS      IL3-UT0117-300301-538-E01 UT0117 Homo sapiens cDNA, mRNA sequence.
DEFINITION      BI063311
ACCESSION      BI063311.1 GI:14470838
VERSION      EST.
KEYWORDS      Homo sapiens (human)
SOURCE      Homo sapiens
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.
REFERENCE      1 (bases 1 to 158)
AUTHORS      Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE      Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
PUBMED      10737800
COMMENT      Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL3&t=IL3-UT0117-
300301-538-E01&t3=2001-03-30&t4=1)
Seq primer: puc 18 forward

FEATURES
source
High quality sequence stop: 158.
Location/Qualifiers
1..158
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="UT0117"
/note="Organ: uterus tumor; Vector: puc18; Site 1: SmaI;
Site 2: SmaI; A mini-library was made by cloning products
derived from QRESPES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

ORIGIN
Query Match 90.4%; Score 30; DB 2; Length 158;
Best Local Similarity 64.3%; Pred. No. 0.049;
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 GCNCAYGCGTGGCCNTCCNTAYAAAGAYTAYGARGTNAAR 42
|||:|||||:|||||:|||||:|||||:|||||:
Db 100 GCCACGCTTGCCGAGCCCTTACAAGACTATGAAGTAAAG 141

RESULT 2
BI063311/c
LOCUS      BE818833      307 bp      mRNA      linear      EST 21-SEP-2000
DEFINITION      CM2-BE0302-050700-256-c01 EN0302 Homo sapiens cDNA, mRNA sequence.
ACCESSION      BE818833
VERSION      BE818833.1 GI:10251067
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.
REFERENCE      1 (bases 1 to 307)
AUTHORS      Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE      Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
PUBMED      10737800
COMMENT      Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM2&t=CM2-BE0302-050
700-256-c01&t3=2000-07-05&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 25
High quality sequence stop: 307.

FEATURES
source
High quality sequence stop: 307.
Location/Qualifiers
1..307
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="EN0302"
/note="Organ: breast normal; Vector: puc18; Site 1: SmaI;
Site 2: SmaI; A mini-library was made by cloning products

```



derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

## ORIGIN

Query Match 90.4%; Score 30; DB 7; Length 307;  
Best Local Similarity 64.3%; Pred. No. 0.056; Mismatches 7; Indels 0; Gaps 0;  
Matches 27; Conservative

QY 1 GCNCAYCGTGGCCNTCNCNTAYAAAGAYTAYGARGTNAAR 42  
|||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
Db 242 GCCCAGCGTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 201

## RESULT 3

BE818817/c  
LOCUS  
DEFINITION CM2-BN0302-050700-256-b07 BN0302 Homo sapiens linear EST 21-SEP-2000  
ACCESSION BE818817  
VERSION BE818817.1 GI:10251051  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

## REFERENCE

1 (bases 1 to 341)  
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

## AUTHORS

Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2=CM2-BN0302-050700-256-b07&tl3=2000-07-05&tl4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 10  
High quality sequence stop: 341.

TITLE  
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

## PUBMED

10737800  
COMMENT  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2=CM2-BN0302-050700-256-b07&tl3=2000-07-05&tl4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 10  
High quality sequence stop: 341.

## FEATURES

## source

1. 341  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/dev\_stage="Adult"  
/clone\_lib="BN0302"  
/notes="Organ: breast normal; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

## ORIGIN

Query Match 90.4%; Score 30; DB 7; Length 341;  
Best Local Similarity 64.3%; Pred. No. 0.058; Mismatches 7; Indels 0; Gaps 0;  
Matches 27; Conservative

QY 1 GCNCAYCGTGGCCNTCNCNTAYAAAGAYTAYGARGTNAAR 42  
|||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
Db 253 GCCCAGCGTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 212

## RESULT 4

BE818835/c  
LOCUS  
DEFINITION CM2-BN0302-050700-256-c11 BN0302 Homo sapiens linear EST 21-SEP-2000  
ACCESSION BE818835  
VERSION BE818835.1 GI:10251069  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

1 (bases 1 to 384)  
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

## REFERENCE

## AUTHORS

Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2=CM2-BN0302-050700-256-c11&tl3=2000-07-05&tl4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 5  
High quality sequence stop: 384.

## FEATURES

1. 384  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/dev\_stage="Adult"  
/clone\_lib="BN0302"  
/notes="Organ: breast normal; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

## ORIGIN

Query Match 90.4%; Score 30; DB 7; Length 384;  
Best Local Similarity 64.3%; Pred. No. 0.059; Mismatches 7; Indels 0; Gaps 0;  
Matches 27; Conservative

QY 1 GCNCAYCGTGGCCNTCNCNTAYAAAGAYTAYGARGTNAAR 42  
|||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
Db 257 GCCCAGCGTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 216

## RESULT 5

BE818828/c  
LOCUS  
DEFINITION CM2-BN0302-050700-256-a05 BN0302 Homo sapiens linear EST 21-SEP-2000

1 (bases 1 to 396)  
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

## REFERENCE

## AUTHORS

Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2=CM2-BN0302-050700-256-c11&tl3=2000-07-05&tl4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 5  
High quality sequence stop: 384.

## FEATURES

1. 384  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/dev\_stage="Adult"  
/clone\_lib="BN0302"  
/notes="Organ: breast normal; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

## ORIGIN

Query Match 90.4%; Score 30; DB 7; Length 384;  
Best Local Similarity 64.3%; Pred. No. 0.059; Mismatches 7; Indels 0; Gaps 0;  
Matches 27; Conservative

QY 1 GCNCAYCGTGGCCNTCNCNTAYAAAGAYTAYGARGTNAAR 42  
|||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
Db 257 GCCCAGCGTTGGCGAGCCCTTACAAAGACTATGAAGTAAAG 216

## RESULT 5

BE818828/c  
LOCUS  
DEFINITION CM2-BN0302-050700-256-a05 BN0302 Homo sapiens linear EST 21-SEP-2000

1 (bases 1 to 396)  
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)



Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=&t2=CM2-BN0302-040  
700-253-b10&t3=2000-07-04&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 4  
High quality sequence stop: 424.

## FEATURES

source  
1..424  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/dev\_stage="Adult"  
/clone\_lib="BN0302"  
/notes="Organ: breast normal; Vector: puc18; Site\_1: SmaI;  
Site\_2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

## ORIGIN

Query Match 90.4%; Score 30; DB 7; Length 424;  
Best Local Similarity 64.3%; Pred. No. 0.06;  
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 GCNCAYCGNTGGCCNTCNCNTAYAAAGAYTAYGAGTNAAR 42  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
DB 250 GCCCAGCTTGGCGGCCCTTACAAGACTATGAAGTAAAG 209

## RESULT 8

BE818807/c  
LOCUS CD672930 449 bp mRNA linear EST 21-SEP-2000  
DEFINITION CM2-BN0302-040700-253-d02 BN0302 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BE818807  
VERSION BE818807.1 GI:10251041  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
1 (bases 1 to 449)  
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.,  
Goldman, G.H., Carvalho, A.P., Matsukuma, A., Baia, G.S., Simpson, D.H.,  
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V.,  
O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
Simpson, A.J.J.  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags

REFERENCE  
AUTHORS

Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=&t2=CM2-BN0302-040

TITLE  
JOURNAL  
PUBMED  
COMMENT

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

## FEATURES

source  
1..449  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/dev\_stage="Adult"  
/clone\_lib="BN0302"  
/notes="Organ: breast normal; Vector: puc18; Site\_1: SmaI;  
Site\_2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

## ORIGIN

Query Match 90.4%; Score 30; DB 7; Length 449;  
Best Local Similarity 64.3%; Pred. No. 0.061;  
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 GCNCAYCGNTGGCCNTCNCNTAYAAAGAYTAYGAGTNAAR 42  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
DB 240 GCCCAGCTTGGCGGCCCTTACAAGACTATGAAGTAAAG 199

## RESULT 9

CD672930 453 bp mRNA linear EST 24-JUN-2003  
DEFINITION fg18a05-y2 Human Iris cDNA (Normalized): fg Homo sapiens cDNA clone  
fg18a05 5', mRNA sequence.

ACCESSION CD672930  
VERSION CD672930.1 GI:32174661  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.

## REFERENCE

1 (bases 1 to 453)  
Wistow, G., Bernstein, S.L., Ray, S., Wyatt, M.K., Behal, A.,  
Touchman, J.W., Bouffard, G., Smith, D. and Peterson, K.  
Expressed sequence tag analysis of adult human iris for the NEIBank  
Project: steroid-response factors and similarities with retinal  
pigment epithelium

JOURNAL  
PUBMED  
COMMENT

Mol. Vis. 8 (4), 185-195 (2002)  
12107412  
Contact: Wistow G  
Section on Molecular Structure and Function  
National Eye Institute  
6/331, NIH, Bethesda, MD 20892-2740, USA  
Tel: 301 402 3452  
Fax: 301 496 0078  
Email: graeme@helix.nih.gov  
Plate: 18 row: a column: 05  
Seq primer: M3RP1 reverse primer (ABI).

## FEATURES

source  
1..453  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="fg18a05"  
/tissue="iris"  
/dev\_stage="Adult"  
/lab\_host="EMDH10B"  
/clone\_lib="Human Iris cDNA (Normalized): fg"  
/notes="Organ: Eye; Vector: pCMVSPOR16; A human iris  
library (bx) was normalized by self-subtraction. One  
portion of double stranded plasmid DNA representing the  
library was linearized by NotI. This NotI digested library

```

/note=Organ: breast_normal; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 136,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
/Clone_ID="ENU302"

```

DB

```

RESULT 12
BE818852/c
LOCUS      467 bp      mRNA      linear      EST 21-SEP-2000
DEFINITION CH2-BN0302-100700-259-a05 BN0302 Homo sapiens cDNA, mRNA sequence.
ACCESSION  BE818852
VERSION     BE818852.1 GI:10251086
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominiidae; Homo.
REFERENCE   1 (bases 1 to 467)
AUTHORS    Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
            O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
TITLE      Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
PUBMED     10737800
COMMENT    Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?ti=&t2=CM2-BN0302-100
            700-259-a05&t3=2000-07-10&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 19
            High quality sequence stop: 467.
            Location/Qualifiers
FEATURES   source
            1..467
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /dev_stage="Adult"
                /clone_lib="BN0302"
                /note="Organ: breast_normal; Vector: puc18; Site_1: SmaI;
                Site_2: SmaI; A mini-library was made by cloning products
                derived from ORESTES PCR (U.S. Letters Patent application
                No. 196,716 - Ludwig Institute for Cancer Research)
                profiles into the pUC 18 vector. Reverse transcription of
                tissue mRNA and cDNA amplification were performed under
                low stringency conditions."
ORIGIN
            Query Match      90.4%; Score 30; DB 7; Length 467;
            Best Local Similarity 64.3%; Pred. No. 0.061;
            Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

            QY 1 GCNCAAGCTTGGCCNTCCNCTAYAAAGAYTAYGARGTNAAR 42
                |||||
                |||||
            Db 246 GCCCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 205
                |||||
                |||||

            RESULT 13
            BP253378
            LOCUS      571 bp      mRNA      linear      EST 15-SEP-2004
            DEFINITION BP253378 Sugano cDNA library, kidney epithelial cell Homo sapiens
            cDNA clone HRC03720, mRNA sequence.
            ACCESSION  BP253378
            VERSION     BP253378.1 GI:52135659
            KEYWORDS    EST.
            SOURCE      Homo sapiens (human)
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominiidae; Homo.
            REFERENCE   1 (bases 1 to 571)
            AUTHORS    Suzuki,Y., Yamashita,R., Shiota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
            TITLE      Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
            JOURNAL    Genome Res. 14 (9), 1711-1718 (2004)
            PUBMED     15342556
            COMMENT    Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Tel: 81-3-5449-5343
            Fax: 81-3-5449-5416
            Email: yusuzuki@hgc.jp.
            Location/Qualifiers
FEATURES   source
            1..571
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone_lib="HRC03720"
                /tissue_type="kidney"
                /cell_type="epithelial cell"
                /clone_lib="Sugano cDNA library, kidney epithelial cell"
ORIGIN
            Query Match      90.4%; Score 30; DB 3; Length 571;
            Best Local Similarity 64.3%; Pred. No. 0.064;
            Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

            QY 1 GCNCAAGCTTGGCCNTCCNCTAYAAAGAYTAYGARGTNAAR 42
                |||||
                |||||
            Db 522 GCCCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 563
                |||||
                |||||

            RESULT 14
            CB152636
            LOCUS      577 bp      mRNA      linear      EST 29-JAN-2003
            DEFINITION K-EST0209868 L12JSHC0s1 Homo sapiens cDNA clone L12JSHC0s1-5-G07
            5', mRNA sequence.
            ACCESSION  CB152636
            VERSION     CB152636.1 GI:28137590
            KEYWORDS    EST.
            SOURCE      Homo sapiens (human)
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominiidae; Homo.
            REFERENCE   1 (bases 1 to 577)
            AUTHORS    Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
            Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
            Kim,Y.S.
            TITLE      21C Frontier Korean EST Project 2001
            JOURNAL    Unpublished (2002)
            COMMENT    Contact: Kim YS
            Genome Research Center
            Korea Research Institute of Bioscience & Biotechnology
            52 Eoeun-dong Yusong-gu, Daejeon 305-333, South Korea
            Tel: +82-42-860-4470
            Fax: +82-42-860-4409
            Email: yongsung@mail.kribb.re.kr
            Plate: 5 row: G column: 07
            High quality sequence stop: 577.
            Location/Qualifiers
FEATURES   source
            1..577
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="L12JSHC0s1-5-G07"
                /sex="M"

```





```

SOURCE      Homo sapiens (human)
ORGANISM
REFERENCE
AUTHORS      Suzuki,Y., Yanashita,R., Shiota,M., Sakakibara,Y., Chiba,J.,
              Mizushima-Sugano,J., Nakai,K. and Sugano,S.
TITLE        Sequence comparison of human and mouse genes reveals a homologous
              block structure in the promoter regions
JOURNAL      Genome Res. 14 (9), 1711-1718 (2004)
PUBMED       15342556
COMMENT      Contact: Yutaka Suzuki
              Department of Virology
              Institute of Medical Science, University of Tokyo
              4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
              Tel: 81-3-5449-5343
              Fax: 81-3-5449-5416
              Email: ysuzuki@hgc.jp.
FEATURES
source
1. .584
   /organism="Homo sapiens"
   /mol_type="mRNA"
   /db_xref="taxon:9606"
   /clone="HRC10676"
   /tissue_type="kidney"
   /cell_type="epithelial cell"
   /clone_lib="Sugano CDNA library, kidney epithelial cell"
ORIGIN
Query Match      90.4%; Score 30; DB 3; Length 584;
Best Local Similarity 64.3%; Pred. No. 0.064;
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;
QY      1 GCNCAYGCGTGGCCNTCNCNTAYAAAGAYTAYGARGTNAAR 42
      |||:||||| |||:||||| |||:||||| |||:|||||
Db      314 GCCCAGCTGGCCGAGCCCTTACAAGACTATGAAGTAAAG 457

RESULT 21
LOCUS      CB130950
DEFINITION K-EST0180946 L12JSHC0 Homo sapiens cDNA clone L12JSHC0-5-D03 5',
mRNA sequence.
ACCESSION  CB130950.1 GI:28095412
VERSION     CB130950
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE   1 (bases 1 to 621)
AUTHORS     Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.S., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.
TITLE       21C Frontier Korean EST Project 2001
JOURNAL     Unpublished (2002)
COMMENT     Contact: Kim YS
              Genome Research Center
              Korea Research Institute of Bioscience & Biotechnology
              52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea
              Tel: +82-42-860-4470
              Fax: +82-42-860-4409
              Email: yongsung@mail.kribb.re.kr
              Plate: 5 row: D column: 03
              High quality sequence stop: 621.
              Location/Qualifiers
                1. .621
                 /organism="Homo sapiens"
                 /mol_type="mRNA"
                 /db_xref="taxon:9606"
                 /clone="L12JSHC0-5-D03"

SOURCE      Homo sapiens (human)
ORGANISM
REFERENCE
AUTHORS      Suzuki,Y., Yanashita,R., Shiota,M., Sakakibara,Y., Chiba,J.,
              Mizushima-Sugano,J., Nakai,K. and Sugano,S.
TITLE        Sequence comparison of human and mouse genes reveals a homologous
              block structure in the promoter regions
JOURNAL      Genome Res. 14 (9), 1711-1718 (2004)
PUBMED       15342556
COMMENT      Contact: Yutaka Suzuki
              Department of Virology
              Institute of Medical Science, University of Tokyo
              4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
              Tel: 81-3-5449-5343
              Fax: 81-3-5449-5416
              Email: ysuzuki@hgc.jp.
FEATURES
source
1. .594
   /organism="Homo sapiens"
   /mol_type="mRNA"
   /db_xref="taxon:9606"
   /clone="HRC10676"
   /tissue_type="kidney"
   /cell_type="epithelial cell"
   /clone_lib="Sugano CDNA library, kidney epithelial cell"
ORIGIN
Query Match      90.4%; Score 30; DB 3; Length 584;
Best Local Similarity 64.3%; Pred. No. 0.064;
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;
QY      1 GCNCAYGCGTGGCCNTCNCNTAYAAAGAYTAYGARGTNAAR 42
      |||:||||| |||:||||| |||:||||| |||:|||||
Db      314 GCCCAGCTGGCCGAGCCCTTACAAGACTATGAAGTAAAG 457

RESULT 22
LOCUS      BE747107
DEFINITION BE747107 630 bp mRNA linear EST 15-SEP-2000
mRNA sequence.
ACCESSION  BE747107.1 GI:10161099
VERSION     BE747107
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE   1 (bases 1 to 630)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
              Email: cgabbs-r@mail.nih.gov
              Tissue Procurement: DCTD/DTF
              CDNA library preparation: Ling Hong/Rubin Laboratory
              CDNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
              Plate: LLCW762 row: o column: 19
              High quality sequence stop: 628.
              Location/Qualifiers
                1. .630
                 /organism="Homo sapiens"
                 /mol_type="mRNA"
                 /db_xref="taxon:9606"
                 /clone="IMAGE:3929706"
                 /tissue_type="adenocarcinoma cell line"
                 /lab_host="DH10B (phage-resistant)"
                 /clone_lib="NIH MGC 9"
                 /note="Organ: ovary; Vector: pOTB7; Site 1: XhoI; Site 2:
                 EcoRI; cDNA made by oligo-dT priming. Directionally
                 cloned into EcoRI/XhoI sites using the following 5'
                 adaptor: GGCACGAG(G). Size-selected >500bp for average
                 insert size 1.8kb. Library constructed by Ling Hong in
                 the laboratory of Gerald M. Rubin (University of

```







Manuscript submitted.

```

ORIGIN
  Query Match          90.4%; Score 30; DB 4; Length 865;
  Best Local Similarity 64.3%; Pred. No. 0.069;
  Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 GCNCAYCGNTGGCCNTCNCNTAYAAAGAYTAYGARGTNAAR 42
   |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
Db 611 GCCCAGCTTGGCGGACCCCTTACAAAGACTATGAAGTAAAG 652

RESULT 28
BM011054          877 bp mRNA linear EST 30-OCT-2001
LOCUS
DEFINITION
603634744F1 NIH_MGC_47 Homo sapiens cDNA clone IMAGE:5421755 5',
mRNA sequence.
ACCESSION
BM011054
VERSION
BM011054.1 GI:16525408
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens

REFERENCE
1 (bases 1 to 877)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
NIH-MGC http://mgs.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM1878 row: h column: 12
High quality sequence stop: 746.
Location/Qualifiers
1 . 877
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5421755"
/tissue_type="neuroblastoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC 47"
/notes="Organ: brain; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

ORIGIN
  Query Match          90.4%; Score 30; DB 2; Length 877;
  Best Local Similarity 64.3%; Pred. No. 0.069;
  Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 GCNCAYCGNTGGCCNTCNCNTAYAAAGAYTAYGARGTNAAR 42
   |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
Db 184 GCCCAGCTTGGCGGACCCCTTACAAAGACTATGAAGTAAAG 225

RESULT 29
BQ958341          899 bp mRNA linear EST 21-AUG-2002
LOCUS
DEFINITION
AGENCOURT_10037048 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6480481

```

```

ACCESSION
BQ958341
VERSION
BQ958341.1 GI:22373819
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens

REFERENCE
1 (bases 1 to 899)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
NIH-MGC http://mgs.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: DCTD/DTF
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM2660 row: f column: 02
High quality sequence stop: 636.
Location/Qualifiers
1 . 899
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6480481"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_40"
/notes="Organ: prostate; Vector: pOTB7; Site 1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

ORIGIN
  Query Match          90.4%; Score 30; DB 3; Length 899;
  Best Local Similarity 64.3%; Pred. No. 0.07;
  Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 GCNCAYCGNTGGCCNTCNCNTAYAAAGAYTAYGARGTNAAR 42
   |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
Db 546 GCCCAGCTTGGCGGACCCCTTACAAAGACTATGAAGTAAAG 587

RESULT 30
BQ831564          901 bp mRNA linear EST 22-MAY-2001
LOCUS
DEFINITION
602765819F1 NIH_MGC_42 Homo sapiens cDNA clone IMAGE:4907885 5',
mRNA sequence.
ACCESSION
BQ831564
VERSION
BQ831564.1 GI:14179151
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens

REFERENCE
1 (bases 1 to 901)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
NIH-MGC http://mgs.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory

```

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: L1CM1813 row: e column: 06  
 High quality sequence stop: 663.  
 Location/Qualifiers

#### FEATURES

1..901  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4907885"  
 /tissue\_type="epithelioid carcinoma cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC 42"  
 /note="Organ: pancreas; Vector: pOTB7; Site 1: XhoI;  
 Site 2: EcoRI; cDNA made by oligo-dT priming.  
 Directionally cloned into EcoRI/XhoI sites using the  
 following 5' adaptor: GGCACGAG(G). Size-selected >500bp  
 for average insert size 1.8kb. Library constructed by Ling  
 Hong in the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies).  
 Note: this is a NIH\_MGC library. |"

#### ORIGIN

Query Match 90.4%; Score 30; DB 2; Length 901;  
 Best Local Similarity 64.3%; Pred. No. 0.07;  
 Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 GNCAYGNTGGCCNTCNCNTAYAARGAYTAYGARGTNAAR 42  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 322 GCCCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 363

#### RESULT 31

BX372160/c  
 LOCUS BX372160 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens  
 DEFINITION cDNA clone CS0DK011YF17 3-PRIME, mRNA sequence.  
 ACCESSION BX372160  
 VERSION BX372160.1 GI:30452082  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

#### REFERENCE

AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage  
 2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
 end enriched, double-strand cDNA was digested with Not I and cloned  
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
 was normalized. Library was constructed by Life Technologies, a  
 division of Invitrogen. This sequence belongs to sequence cluster  
 6601.f

For more information about this cluster, see  
[http://www.genoscope.cns.fr/cdna?s=CS0BA1040ZE07\\_CS03797\\_1&c=6601.f](http://www.genoscope.cns.fr/cdna?s=CS0BA1040ZE07_CS03797_1&c=6601.f)

#### FEATURES

Location/Qualifiers  
 1..918  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0DK011YF17"  
 /cell\_type="HELA CELLS COT 25-NORMALIZED"

/cell\_line="HELA"  
 /clone\_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-oligo(dT)  
 primer. Five prime end enriched, double-strand cDNA was  
 digested with Not I and cloned into the Not I and EcoR V  
 sites of the pCMVSPORT 6 vector. Library was normalized."

#### ORIGIN

Query Match 90.4%; Score 30; DB 4; Length 918;  
 Best Local Similarity 64.3%; Pred. No. 0.07;  
 Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 GNCAYGNTGGCCNTCNCNTAYAARGAYTAYGARGTNAAR 42  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 335 GCCCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 294

#### RESULT 32

BUS43952  
 LOCUS BUS43952 923 bp mRNA linear EST 13-SEP-2002  
 DEFINITION AGENCOURT\_10333041 NIH\_MGC\_40 Homo sapiens cDNA clone IMAGE:6576152  
 5', mRNA sequence.  
 ACCESSION BUS43952  
 VERSION BUS43952.1 GI:22854435  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Hominidae; Homo.  
 1 (bases 1 to 923)  
 NIH-MGC <http://mgc.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 JOURNAL  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: DCTB/DTP

#### REFERENCE

CDNA Library Preparation: Rubin Laboratory  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: L1CM2774 row: d column: 08  
 High quality sequence stop: 641.  
 Location/Qualifiers

#### FEATURES

1..923  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6576152"  
 /tissue\_type="carcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC 40"  
 /note="Organ: prostate; Vector: pOTB7; Site 1: XhoI;  
 Site 2: EcoRI; cDNA made by oligo-dT priming.  
 Directionally cloned into EcoRI/XhoI sites using the  
 following 5' adaptor: GGCACGAG(G). Library constructed by  
 Ling Hong in the laboratory of Gerald M. Rubin (University  
 of California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies).  
 Note: this is a NIH\_MGC library."

#### ORIGIN

Query Match 90.4%; Score 30; DB 3; Length 923;  
 Best Local Similarity 64.3%; Pred. No. 0.07;  
 Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 GNCAYGNTGGCCNTCNCNTAYAARGAYTAYGARGTNAAR 42  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 546 GCCCAGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 587

```
RESULT 33
BX370558
LOCUS
DEFINITION BX370558 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens
cDNA clone CS0DK011YF17 5-PRIME, mRNA sequence.
ACCESSION BX370558
VERSION BX370558.1 GI:30453927
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 962)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
6601.f
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?s=CS0BAG0492C03_CS04627_1&c=6601.f

FEATURES
source
1. .962
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DK011YF17"
/cell_type="HELA CELLS COT 25-NORMALIZED"
/cell_line="HELA"
/clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and EcoR V sites of the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
Query Match 90.4%; Score 30; DB 4; Length 962;
Best Local Similarity 64.3%; Pred. No. 0.071;
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 GCNCAYGCNTGGCCNTCCNTAYAAAGAYTAYGARGTNAAR 42
Db 248 GCCACGCTTGGCCGAGCCCTTACAAGACTATGAAGTAAAG 289
|||||
|||||

RESULT 34
BQ067433
LOCUS
DEFINITION BQ067433 1023 bp mRNA linear EST 02-APR-2002
AGENCOURT_6758944 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5754944
5', mRNA sequence.
ACCESSION BQ067433
VERSION BQ067433.1 GI:19896479
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 1023)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov

FEATURES
source
1. .1830
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"

Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
plate: LLAMI2793 row: C column: 09
High quality sequence stop: 649.

FEATURES
source
1. .1023
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5754944"
/lab_host="DH10B"
/clone_lib="NIH_MGC_115"
/notes="Organ: pooled brain, lung, testis; Vector:
pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA
source anonymous pool of 6 male brains, age range 23-27; 1
male lung, age 27; and 1 male testis, age 69. Library is
oligo-dT primed and directionally cloned (EcoRV site is
destroyed upon cloning). Average insert size 1.8 kb,
insert size range 1-3 kb. Library is normalized and
enriched for full-length clones and was constructed by C.
Gruber (Invitrogen). Research Genetics tracking code
021. Note: this is a NIH_MGC Library."

ORIGIN
Query Match 90.4%; Score 30; DB 3; Length 1023;
Best Local Similarity 64.3%; Pred. No. 0.072;
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 GCNCAYGCNTGGCCNTCCNTAYAAAGAYTAYGARGTNAAR 42
Db 403 GCCACGCTTGGCCGAGCCCTTACAAGACTATGAAGTAAAG 444
|||||
|||||

RESULT 35
CR597125
LOCUS
DEFINITION CR597125 1830 bp mRNA linear HTC 21-JUL-2004
full-length cDNA clone CS0DK011YF17 of HeLa cells COT 25-normalized
of Homo sapiens (human).
ACCESSION CR597125
VERSION CR597125.1 GI:50477932
KEYWORDS HTC; CNSiUT_cDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 1830)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
Contact: Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue
2 (bases 1 to 1830)
Genoscope.
Direct Submission
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.

FEATURES
source
1. .1830
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
```

```

/db_xref="taxon:9606"
/clone="CS0BK011YF17"
/tissue_type="HeLa cells Cot 25-normalized"
/plasmid="PCWVSPORT_6"

ORIGIN
Query Match          90.4%; Score 30; DB 6; Length 1830;
Best Local Similarity 64.3%; Pred. No. 0.08;
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 GCNCAYGNTGGCCNTCNCNTAYAAAGAYTAYGAGTNAAR 42
    |||:||||| |||:|||||:|||||:|||||:
Db 1118 GCCACGCTTGGCCGAGCCCTTACAAGACTATGAGTAAAG 1159

RESULT 36
BI193620
LOCUS
DEFINITION
602946519F1 NIH_MGC_42 Homo sapiens cDNA clone IMAGE:5089900 5',
mRNA sequence.
ACCESSION
BI193620
VERSION
BI193620.1 GI:14648640
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1 (bases 1 to 1125)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1851 row: e column: 05
High quality sequence stop: 529.
Location/Qualifiers
1..1125
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5089900"
/tissue_type="epithelioid carcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 42"
Note: "Organ: pancreas; Vector: pOTB7; Site 1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library. |"

ORIGIN
Query Match          88.6%; Score 29.4; DB 2; Length 1125;
Best Local Similarity 65.9%; Pred. No. 0.13;
Matches 27; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 1 GCNCAYGNTGGCCNTCNCNTAYAAAGAYTAYGAGTNAAA 41
    |||:||||| |||:|||||:|||||:|||||:
Db 294 GCCACGCTTGGCCGAGCCCTTACAAGACTATGAAAGTAAA 334

RESULT 37

```

```

CF780868
LOCUS
DEFINITION
AGENCOURT_15939245 NIH_MGC_219 Homo sapiens cDNA clone
IMAGE:30523569 5', mRNA sequence.
ACCESSION
CF780868
VERSION
CF780868.1 GI:37740645
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1 (bases 1 to 916)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: James Martin, University of Iowa
CDNA Library Preparation: M. Bento Soares, University of Iowa
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: NDAM601 row: 1 column: 10
High quality sequence stop: 651.
Location/Qualifiers
1..916
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30523569"
/tissue_type="Pooled Chondrosarcoma Tumor cells"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC_219"
Note: "Vector: pYX-Asc; Site 1: EcoRI; Site 2: NotI;
Library is oligo-dT primed and directionally cloned
Denatured RNA was size fractionated on a 1% agarose gel.
First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated with
EcoR I adaptor, digested with Not I and then cloned
directionally into pYX-Asc vector. Average insert size
0.5-1kb. Adaptors 5'(AATTCGACGAGG)3' and 5'd
(CCTCGTGGCG)3'. 3' Linker sequence - GCGGCGCTGAGAGCC T18.
Sequencing primers 3'end: T3 promoter primer 5'd
(AATTACCTCTACTAAGGGA)3'. 5' End: T7 promoter primer 5'd
(TAATACGACTCACTATAGG)3'. Library was constructed in the
laboratory of M. Bento Soares. Average insert size 2-3kb.
Note: this is a NIH_MGC Library."

ORIGIN
Query Match          87.3%; Score 29; DB 5; Length 916;
Best Local Similarity 61.9%; Pred. No. 0.19;
Matches 26; Conservative 7; Mismatches 9; Indels 0; Gaps 0;

QY 1 GCNCAYGNTGGCCNTCNCNTAYAAAGAYTAYGAGTNAAR 42
    |||:||||| |||:|||||:|||||:|||||:
Db 667 GCCACGCTTGGCCGAGCCCTTACAAGACTATGAAAGTAAAG 708

RESULT 38
BE18831/c
LOCUS
DEFINITION
CM2-BN0302-050700-256-a12 BN0302 Homo sapiens cDNA, mRNA sequence.
ACCESSION
BE18831
VERSION
BE18831.1 GI:10251065
KEYWORDS
EST.
SOURCE
Homo sapiens (human)

```

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.

REFERENCE  
AUTHORS 1 (bases 1 to 365)  
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,  
Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V.,  
O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
Simpson, A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags

JOURNAL  
PUBLISHED Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2=CM2-BN0302-050  
700-256-al2&t3=2000-07-05&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 14  
High quality sequence stop: 365.

FEATURES  
source  
1..365  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/dev\_stage="Adult"  
/clone\_lib="BN0302"  
/notes="Organ: breast normal; Vector: puc18; Site: 1: SmaI;  
Site 2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

ORIGIN  
Query Match 85.5%; Score 28.4; DB 7; Length 365;  
Best Local Similarity 61.9%; Pred. No. 0.3;  
Matches 26; Conservative 7; Mismatches 9; Indels 0; Gaps 0;  
Qy 1 GCNCAYGCVTGGCCNTCCNTAYAAAGATYAGAGTNAAR 42  
|||:||||| |||:|||||:|||||:|||||:  
Db 267 GCCACGCTGGCGGCGCCCTTACAAAGACTATGAAGTAAAG 226

RESULT 39  
BQ685729  
LOCUS BQ685729 940 bp mRNA linear EST 15-JUL-2002  
DEFINITION AGENCOURT\_8344399 NIH\_MGC\_110 Homo sapiens cDNA clone IMAGE:6251057  
5', mRNA sequence.

ACCESSION BQ685729  
VERSION BQ685729.1 GI:21811045  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.

REFERENCE  
AUTHORS 1 (bases 1 to 940)  
NIH-MGC http://mgi.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.

ORGANISM Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.

REFERENCE  
AUTHORS 1 (bases 1 to 365)  
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,  
Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V.,  
O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
Simpson, A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags

JOURNAL  
PUBLISHED Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2=CM2-BN0302-050  
700-256-al2&t3=2000-07-05&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 14  
High quality sequence stop: 365.

FEATURES  
source  
1..365  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/dev\_stage="Adult"  
/clone\_lib="BN0302"  
/notes="Organ: breast normal; Vector: puc18; Site: 1: SmaI;  
Site 2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

ORIGIN  
Query Match 85.5%; Score 28.4; DB 7; Length 365;  
Best Local Similarity 61.9%; Pred. No. 0.3;  
Matches 26; Conservative 7; Mismatches 9; Indels 0; Gaps 0;  
Qy 1 GCNCAYGCVTGGCCNTCCNTAYAAAGATYAGAGTNAAR 42  
|||:||||| |||:|||||:|||||:|||||:  
Db 267 GCCACGCTGGCGGCGCCCTTACAAAGACTATGAAGTAAAG 226

RESULT 39  
BQ685729  
LOCUS BQ685729 940 bp mRNA linear EST 15-JUL-2002  
DEFINITION AGENCOURT\_8344399 NIH\_MGC\_110 Homo sapiens cDNA clone IMAGE:6251057  
5', mRNA sequence.

ACCESSION BQ685729  
VERSION BQ685729.1 GI:21811045  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.

REFERENCE  
AUTHORS 1 (bases 1 to 940)  
NIH-MGC http://mgi.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgabs-x@mail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Rubin Laboratory  
DNA sequencing by: The I.M.A.G.E. Consortium (LLNL)  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLC2395 row: b column: 18  
High quality sequence stop: 538.

FEATURES  
source  
1..940  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6251057"  
/tissue\_type="ductal carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_110"  
/notes="Organ: pancreas; Vector: pOTB7; Site: 1: XhoI;  
Site 2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCACGAG(G). Library constructed by  
Ling Hong in the laboratory of Gerald M. Rubin (University  
of California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library."

ORIGIN  
Query Match 85.5%; Score 28.4; DB 3; Length 940;  
Best Local Similarity 61.9%; Pred. No. 0.36;  
Matches 26; Conservative 7; Mismatches 9; Indels 0; Gaps 0;  
Qy 1 GCNCAYGCVTGGCCNTCCNTAYAAAGATYAGAGTNAAR 42  
|||:||||| |||:|||||:|||||:|||||:  
Db 687 GCCACGCTGGCGGCGCCCTTACAAAGACTATGAAGTAAAG 728

RESULT 40  
AUI40676  
LOCUS AUI40676 680 bp mRNA linear EST 05-AUG-2002  
DEFINITION AUI40676 PLACE4 Homo sapiens cDNA clone PLACE4000034 5', mRNA  
sequence.

ACCESSION AUI40676  
VERSION AUI40676.1 GI:11002197  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.

REFERENCE  
AUTHORS 1 (bases 1 to 680)  
Kimura, K., Wakamatsu, A., Suzuki, Y., Ota, T., Nishikawa, T.,  
Yamashita, R., Yamamoto, J., Sekine, M., Tsuritani, K., Wakaguri, H.,  
Ishii, S., Sugiyama, T., Saito, K., Isono, Y., Irie, R., Kuehida, N.,  
Yoneyama, T., Otsuka, R., Kanda, K., Yokoi, T., Kondo, H., Wagatsuma, M.,  
Murakawa, K., Ishida, S., Ishibashi, T., Takahashi, Fujii, A.,  
Tanase, T., Nagai, K., Kikuchi, H., Nakai, K., Isogai, T. and Sugano, S.  
Diversification of Transcriptional Modulation: Large-scale  
Identification and Characterization of Putative Alternative  
Promoters of Human Genes  
Genome Res. 16 (1), 55-65 (2006)  
16344560  
Contact: Takao Isogai  
FLJ Project (HRI Team)  
Helix Research Institute  
2-6-7 Kazusa-Kamatari, Kisarazu, Chiba, 292-0818, Japan  
Tel: 81-438-52-3975  
Fax: 81-438-52-3986  
Email: flj-cdna@nifty.com  
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix  
Research Institute; cDNA library construction: Department of  
Virology, Institute of Medical Science, University of Tokyo, and

```

FEATURES
source
Helix Research Institute.
Location/Qualifiers
1. .680
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="PLACE4000034"
/tissue_type="placenta"
/clone_lib="PLACE4"
/note="Vector: pME189FL3"

ORIGIN
Query Match      82.5%; Score 27.4; DB 1; Length 680;
Best Local Similarity 59.5%; Pred. No. 0.93;
Matches 25; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

QY 1 GCNCAAGTCGCTCCNTCNCNTAYAARGAYTAYGARTNAAR 42
Db 633 GCCACGGCTTGCGCCCAACCCTTACAAGACTATGAANGTA 674

RESULT 41
AA615518          454 bp mRNA linear EST 07-OCT-1997
LOCUS vo7oe02.r1 Barstead mouse myotubes MPLRB5 Mus musculus cDNA clone
DEFINITION IMAGE:1064474 5', mRNA sequence.
ACCESSION AA615518
VERSION AA615518
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE
1 (bases 1 to 454)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
The WashU-HMMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HMMI Mouse EST Project
Washington University School of Medicine#
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:586834
Seq primer: -28ml3 rev2 ET from Amer sham.

FEATURES
source
Location/Qualifiers
1. .454
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C3H"
/db_xref="taxon:10090"
/clone="IMAGE:1064474"
/cell_line="C2C12"
/lab_host="DH10B"
/clone_lib="Barstead mouse myotubes MPLRB5"
/note="Vector: pPTD3D-PacI; Site1: EcoRI; Site 2: NotI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACGAATCTGAAGTGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors [AATTCCGATCCTTC], digested with Not I and cloned into the Not I and Eco RI sites of the modified pRT3 vector. Library constructed by Bob Barstead. The C2C12 cell line (available from ATCC, catalog # CRL-1772) differentiates rapidly, forming contractile myotubes and producing

FEATURES
source
Location/Qualifiers
1. .652
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="mRNA"
/cultivar="IR64"
/db_xref="taxon:39946"
/clone="BR060012A20C08.ab1"
/tissue_type="Panicles"
/dev_stages="Flowering"
/clone_lib="IRRI clones"
/note="vector: phuescript II SK+; Water stress was applied by not watering for 4 consecutive days. Panicles were collected from control (well watered) and stressed plants at 2 days before heading, heading, 50% flowering and 4 days after 50% flowering."

ORIGIN
Query Match      71.1%; Score 23.6; DB 4; Length 652;
Best Local Similarity 59.0%; Pred. No. 44;
Matches 23; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY 4 CAYCNTGGCCNTCNCNTAYAARGAYTAYGARTNAAR 42
Db 426 CACGCTTTGCCTTCCCCTCATCTAATGTATGAGCCAAA 464

RESULT 43
CNS02R3E          952 bp DNA linear GSS 01-SEP-2000
LOCUS Tetraodon nigroviridis genome survey sequence T7 end of clone
DEFINITION 159K05 of library G from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION AL210083
VERSION AL210083.1 GI:7868902
KEYWORDS GSS; genome survey sequence.
```







Db 1058 CCCACGGCTTGGCCGTCTCCCTGTTTAAGACCGCGGACAA 1097

## RESULT 48

DT989670

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CONTACT: Grimwood, Jane

Stanford Human Genome Center

Stanford University School of Medicine

975 S California Ave, Palo Alto, CA 94304, USA

Tel: 650 320 5917

Fax: 650 320 5801

Email: jane@shgc.stanford.edu

Plate: 245

High quality sequence stop: 592.

Location/Qualifiers

1..1439

/organism="Gasterosteus aculeatus"

/mol\_type="mRNA"

/strain="Bittorfjordur marine sticklebacks, Iceland"

/db\_xref="taxon:69293"

/clones="CLJ245-B03"

/sex="mixed male and female"

/tissue\_type="whole larva"

/dev\_stage="21 day old larvae collected at Swarup Stage 30 (J. Embryol. Exp. Morphol 6: 373-383.1958)"

/lab\_host="DH10B (T1 phage resistant)"

/clone\_lib="SHGC-CLJ"

/notes="Vector: Express 1; Total and poly A+ RNA was isolated from the indicated stickleback tissue, and a cDNA library was constructed in the Express 1 plasmid vector by Open Biosystems. First strand cDNA synthesis was primed with an 54 bp linker containing an oligodT sequence preceded by a synthetic NotI site (first strand primer: 5'-GACTAGTCTAGATCGGAGCGGCC(T)25-3'). Following second strand synthesis, cDNAs were made blunt at the end corresponding to the original 5 prime end of mRNA, and cloned directionally into the NotI and EcoRV sites of Express 1. Note that the EcoRV site is typically destroyed in the blunt end cloning, leaving a junction of the form 'xxxATC' (where is ATC is the second half of the EcoRV site, and xxx is derived from the cDNA sequence). A map of the Express 1 vector is available at: http://www.openbiosystems.com/cdna\_library\_construction\_fa q.php#8 The primary library was transformed and amplified in DH10B (T1 phage resistant) bacteria. Clones available from Open Biosystems: http://www.openbiosystems.com/stickleback"

Query Match 67.5%; Score 22.4; DB 10; Length 1439;

Best Local Similarity 57.5%; Pred. No. 1.8e+02;

Matches 23; Conservative 5; Mismatches 12; Indels 0; Gaps 0;

Qy 2 CNAAYGNTGCCNTCCNTAAYARGAYTAARGTNAA 41

Db 133 CTCACGATGGCTTCCCGTATGAAGATGATCCTGTGAA 172

## ORIGIN

## RESULT 49

BE006062

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2-RC2-BN0123-170 300-011-c07&t3=2000-03-17&t4=1)

Seq primer: puc 18 forward

High quality sequence start: 24

High quality sequence stop: 583.

Location/Qualifiers

1..583

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/dev\_stage="Adult"

/clone\_lib="BN0123"

/notes="Organ: breast\_normal; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

Query Match 66.9%; Score 22.2; DB 7; Length 583;

Best Local Similarity 65.6%; Pred. No. 1.8e+02;

Matches 21; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

Qy 1 GCNAYGNTGCCNTCCNTAAYARGAYTA 32

Db 552 GCCACGCTTGGCGGCGGCTTACAAAGACTA 583

## RESULT 50

AQ692009/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

Qy 1 GCMAYGNTGCCNTCCNTAAYARGAYTA 32

Db 552 GCCACGCTTGGCGGCGGCTTACAAAGACTA 583

## ORIGIN

Query Match 67.5%; Score 22.4; DB 10; Length 1439;

Best Local Similarity 57.5%; Pred. No. 1.8e+02;

Matches 23; Conservative 5; Mismatches 12; Indels 0; Gaps 0;

Qy 2 CNAAYGNTGCCNTCCNTAAYARGAYTAARGTNAA 41

Db 133 CTCACGATGGCTTCCCGTATGAAGATGATCCTGTGAA 172

**SOURCE** Homo sapiens (human)  
**ORGANISM** Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.  
**REFERENCE** 1 (bases 1 to 560)  
**AUTHORS** Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.  
**TITLE** Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome  
**JOURNAL** Proc.Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)  
**PUBMED** 1049764  
**COMMENT** Contact: Mahairas GG, Wallace JC, Hood L  
 High Throughput Sequencing Center  
 University of Washington  
 401 Queen Anne Avenue North, Seattle, WA 98109, USA  
 Tel: (206) 616-3618  
 Fax: (206) 616-3887  
 Email: jwallace@u.washington.edu  
 Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering\_bac.htm) or from Resear h Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu  
 Plate: 978 row: P column: 23  
 Seq primer: T7  
 Class: BAC ends  
 High quality sequence stop: 560.  
**FEATURES**  
 source  
 1..560  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /clone="Plate=978 Col=23 Row=P"  
 /sex="male"  
 /clone\_lib="RPCI-11 Human Male BAC Library"  
 /notes="Vector: pBACE3.6; Site 1: EcoRI; Site 2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at EcoRI sites"  
**ORIGIN**  
 Query Match 66.3%; Score 22; DB 11; Length 560;  
 Best Local Similarity 52.4%; Pred. No. 2.2e+02;  
 Matches 22; Conservative 7; Mismatches 13; Indels 0; Gaps 0;  
 QY 1 GCNCGCAGTGGCCNTCCNTAYARGAYTAYGAGTNAAR 42  
 |||||  
 Db 198 GCCCATTCATGGACATCTCCACACAAAGTCTACGGACTTTAA 157  
 |||||  
**RESULT** 51  
**LOCUS** CL815091  
**DEFINITION** OR\_CBa0031004.r OR\_CBa Oryza rufipogon genomic clone OR\_CBa0031004  
 3'- genomic survey sequence.  
**ACCESSION** CL815091  
**VERSION** CL815091.1 GI:51058756  
**KEYWORDS** GSS.  
**SOURCE** Oryza rufipogon  
**ORGANISM** Oryza rufipogon  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP clade; Ehrhartoideae; Oryzeae; Oryza.  
**REFERENCE** 1 (bases 1 to 805)  
**AUTHORS** Kim,H., Yu,Y., Wisseotsaki,M., Yost,D., Stum,D., Rao,K., Luo,M., Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.  
**TITLE** ONAP project  
**JOURNAL** Unpublished (2004)

**COMMENT** Contact: Rod A. Wing  
 Arizona Genomics Institute  
 University of Arizona  
 Forbes Building Room 303, Tucson, AZ 85721-0036, USA  
 Tel: 520 626 9595  
 Fax: 520 621 1259  
 Email: rwing@genome.arizona.edu  
 PCR Primers  
 FORWARD: TAA TAC GAC TCA CTA TAG GG  
 BACKWARD: CAC TCA TTA GGC ACC CCA  
 Plate: 0031 row: O column: 04  
 Seq primer: CAC TCA TTA GGC ACC CCA  
 Class: BAC ends.  
**FEATURES**  
 Location/Qualifiers  
 1..805  
 /organism="Oryza rufipogon"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:4529"  
 /clone="OR\_CBa0031004"  
 /tissue\_type="young leaves"  
 /dev\_stage="2 week old seedlings"  
 /lab\_host="DH10B T1 phage resistant"  
 /clone\_lib="OR\_CBa"  
 /note="Vector: pAG1BAC1; Site 1: HindIII; Site 2: HindIII; drk treated 36 hrs before harvest"  
**ORIGIN**  
 Query Match 66.3%; Score 22; DB 13; Length 805;  
 Best Local Similarity 56.4%; Pred. No. 2.4e+02;  
 Matches 22; Conservative 6; Mismatches 11; Indels 0; Gaps 0;  
 QY 4 CAYGNTGGCCNTCCNTAYARGAYTAYGAGTNAAR 42  
 |||||  
 Db 84 CAGCTTTGCTTCCCTCCCATCTAGTGTATGAGCCAAA 122  
 |||||  
**RESULT** 52  
**AG837338/c**  
**LOCUS** AG837338  
**DEFINITION** Oryza sativa (indica cultivar-group) genomic DNA, BAC end sequence, BAC clone:K0038B07\_R, genomic survey sequence.  
**ACCESSION** AG837338  
**VERSION** AG837338.1 GI:55303573  
**KEYWORDS** GSS.  
**SOURCE** Oryza sativa (indica cultivar-group)  
**ORGANISM** Oryza sativa (indica cultivar-group)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP clade; Ehrhartoideae; Oryzeae; Oryza.  
**REFERENCE** 1  
**AUTHORS** Katagiri,S., Wu,J., Ito,Y., Karasawa,W., Shibata,M., Kanamori,H., Katayose,Y., Namiki,N., Matsumoto,T. and Sasaki,T.  
**TITLE** End Sequencing and Chromosomal in silico Mapping of BAC Clones Derived from an indica Rice Cultivar, Kasalath  
**JOURNAL** Breeding Science 54, 273-279 (2004)  
**REFERENCE** 2 (bases 1 to 842)  
**AUTHORS** Sasaki,T., Matsumoto,T. and Wu,J.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (29-OCT-2004) Takuji Sasaki, National Institute of Agricultural Sciences, Rice Genome Research Program; Kannondai 2-1-2, Tsukuba, Ibaraki, 305-8602, Japan  
 (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/, Tel:81-298-38-7441, Fax:81-298-38-7468)  
**COMMENT** The orientation of the sequence is from SP6 side of the BAC clone.  
**FEATURES**  
 Location/Qualifiers  
 1..842  
 /organism="Oryza sativa (indica cultivar-group)"  
 /mol\_type="genomic DNA"  
 /cultivar="Kasalath"  
 /db\_xref="taxon:39946"  
 /clone="K0038B07\_R"  
**ORIGIN**

```

Query Match      66.3%; Score 22; DB 14; Length 842;
Best Local Similarity 56.4%; Pred. No. 2.4e+02;
Matches 22; Conservative 6; Mismatches 11; Indels 0; Gaps 0;

Qy 4 CAYCGTGGCCNTCNCNTAYARGATAYARGATNAAR 42
Db 126 CACGCTTGGCTTCCCATCTAGTGGTTATGAAGCCAAA 88

RESULT 53
LOCUS CV968179/c 880 bp mRNA linear EST 25-JAN-2005
DEFINITION PC057H02 infected tomato, center of lesion 3 dpi Phytophthora
infestans cDNA, mRNA sequence.
ACCESSION CV968179
VERSION 1 GI:58157928
KEYWORDS EST.
SOURCE Phytophthora infestans (potato late blight agent)
ORGANISM Phytophthora infestans
REFERENCE 1 (bases 1 to 880)
AUTHORS Randall,T., Dwyer,R.A., Huitema,E., Beyer,K., Cvitanich,C.,
Kelkar,H., Fong,A.M., Gates,K., Roberts,S., Yatzkan,E., Gaffney,T.,
Lew,M., Testa,A., Torco-Alalibo,A., Zhang,M., Zheng,L., Mueller,E.,
Windass,J., Binder,A., Birch,P.R.J., Gisi,U., Govers,F., Gow,N.A.,
Mauch,F., van West,P., Waugh,M.E., Yu,J., Boller,T., Kamoun,S.,
Lam,S.T. and Judelson, H.S.
TITLE Large-scale gene discovery in the oomycete Phytophthora infestans
reveals likely components of phytopathogenicity shared with true
fungi
JOURNAL Mol. Plant-Microbe Interact. 18 (3), 229-243 (2005)
PUBLISHED 15782637
COMMENT Contact: Judelson HS
Department of Plant Pathology
University of California
Webster Hall, Riverside, CA 92521, USA
Tel: 909 787 4199
Fax: 909 787 4294
Email: howard.judelson@ucr.edu
may be of host plant origin.
FEATURES
source Location/Qualifiers
1..880
/mol_type="mRNA"
/strain="88069"
/db_xref="taxon:4787"
/sex="Al"
/clone_lib="infected tomato, center of lesion 3 dpi"
/notes="Vector: pSPORT1"
ORIGIN
Query Match      66.3%; Score 22; DB 8; Length 880;
Best Local Similarity 52.4%; Pred. No. 2.4e+02;
Matches 22; Conservative 7; Mismatches 13; Indels 0; Gaps 0;

Qy 1 GCNAYCGTGGCCNTCNCNTAYARGATAYARGATNAAR 42
Db 684 GCCAATAGTGGCCCTCACCTATCAAGTTTGGGGTCGAG 643

RESULT 54
LOCUS CD190259/c 315 bp mRNA linear EST 14-SEP-2003
DEFINITION MS1-0063U-A372-E08-U.G MS1-0063 Schistosoma mansoni cDNA clone
MS1-0063U-A372-E08-G, mRNA sequence.
ACCESSION CD190259
VERSION 1 GI:34720234
KEYWORDS EST.
SOURCE Schistosoma mansoni
ORGANISM Schistosoma mansoni
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
Strigeidida; Schistosomatoidae; Schistosomatidae; Schistosoma.
REFERENCE 1 (bases 1 to 315)

Verjovski-Almeida,S., DeMarco,R., Martins,E.A.L., Guimaraes,P.E.M.,
Ojopi,E.P.B., Paquola,A.C.M., Piazza,J.P., Nishiyama,M.Y. Jr.,
Kitajima,J.P., Adameson,R.E., Ashton,P.D., Bonaldo,M.P.,
Coulson,P.S., Dillon,G.P., Farias,L.P., Gregorio,S.P., Ho,P.L.,
Leite,R.A., Malaquias,L.C.C., Marques,R.C.P., Miyasato,P.A.,
Nascimento,A.L.T.O., Ohlweiler,F.P., Reis,E.M., Ribeiro,M.A.,
Sa,R.G., Stukart,G.C., Soares,M.B., Gargioni,C., Kawano,T.,
Rodrigues,V., Madeira,A.M.B.N., Wilson,R.A., Menck,C.F.M.,
Setubal,J.C., Leite,L.C.C. and Dias-Neto,E.
TITLE Transcription analysis of the acoelomate human parasite Schistosoma
mansoni
JOURNAL Nat. Genet. 35 (2), 148-157 (2003)
PUBLISHED 12973350
COMMENT Contact: Dr. Sergio Verjovski-Almeida
Departamento de Bioquímica
Instituto de Química - Universidade de São Paulo
Av. Prof. Lineu Prestes 748 sala 1200, 05508-900 São Paulo - SP,
Brasil
Tel: +55-11-3091-2173
Fax: +55-11-3091-2186
Email: verjo@iq.usp.br
This sequence was derived from the FAPESP Schistosoma mansoni EST
Genome Project. All sequences in the project were assembled and
annotated. This entry and all the assembled sequences can be seen
in the following URL http://bioinfo.iq.usp.br/schisto/
Plate: MS1-0063U-A372 row: 8 column: E.
FEATURES
source Location/Qualifiers
1..315
/mol_type="Schistosoma mansoni"
/db_xref="taxon:6183"
/clone="MS1-0063U-A372-E08-G"
/sex="mixed pool"
/dev_stage="schistosomulum"
/lab_host="in vitro culture"
/clone_lib="MS1-0063"
/notes="Vector: pGEM T-easy"
ORIGIN
Query Match      65.7%; Score 21.8; DB 5; Length 315;
Best Local Similarity 60.5%; Pred. No. 2.4e+02;
Matches 23; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

Qy 1 GCNAYCGTGGCCNTCNCNTAYARGATAYARGAT 38
Db 58 GCATGATGCGCCAGGACCATCATGATGATGATGT 21

RESULT 55
LOCUS DU888204 936 bp DNA linear GSS 03-JAN-2006
DEFINITION 389797 Tomato HindIII BAC Library Lycopersicon esculentum genomic
clone LE HBa0211C12 3, genomic survey sequence.
ACCESSION DU888204
VERSION 1 GI:84183949
KEYWORDS GSS.
SOURCE Lycopersicon esculentum (Solanum lycopersicum)
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE 1 (bases 1 to 936)
AUTHORS Mueller,L.A., Buels,R.M., Wang,Y., Tanksley,S.D., Giovannoni,J.J.,
Van Eck,J. and Stack,S.
TITLE BAC end sequencing from three Solanum lycopersicon libraries
JOURNAL Unpublished (2005)
COMMENT Other GSSs: 389798
Contact: Lukas Mueller
Tanksley Lab, Dept. of Plant Breeding
Cornell University
251 Emerson Hall, Ithaca, NY 14853, USA
Tel: 607-255-6557
Fax: 607-255-6683

```

```

Email: sgn-feedback@sgn.cornell.edu
Insert length: 22763 Std Error: 0.00
Plate: 211 row: C column: 12
Seq primer: SP6
Class: BAC ends
High quality sequence start: 27
High quality sequence stop: 822.

FEATURES
    source
        1..936
            /organism="Lycopersicon esculentum"
            /mol_type="genomic DNA"
            /cultivar="Heinz 1706"
            /db_xref="taxon:4081"
            /clone="LE_HBA0211C12"
            /lab_host="E. coli"
            /clone_lib="Tomato HindIII BAC Library"
            /notes="Vector: pBelOBAC11, Site_1: HindIII"

ORIGIN
Query Match      65.7%; Score 21.8; DB 14; Length 936;
Best Local Similarity 60.6%; Pred. No. 3e+02;
Matches 20; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

QY 4 CAYGNTGGCCNTCCNTAYAARGAYTAYGAR 36
Db 549 CACTCATGTCATTCCTCTACAAAGACTATGAA 581

RESULT 56
CE724340/c 626 bp DNA linear GSS 30-SEP-2003
LOCUS
DEFINITION
    tigr-gss-dog-1700031521927 Dog Library Canis familiaris genomic,
    genomic survey sequence.
ACCESSION
    CE724340
VERSION
    CE724340.1 GI:37064365
KEYWORDS
    GSS:
SOURCE
    Canis familiaris (dog)
    ORGANISM
        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
        Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
        Canis.
REFERENCE
    1 (bases 1 to 626)
    Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
    Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
    Venter,J.C.
    The dog genome: survey sequencing and comparative analysis
    Science 301 (5641), 1898-1903 (2003)
    PUBMED
    14512627
    Contact: Kirkness EF
    The Institute for Genomic Research
    Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
    Rockville, MD 20850, USA
    Tel: 301-838-0200
    Fax: 301-838-0208
    Email: ekirknes@tigr.org
    Class: shotgun.
    Location/Qualifiers
        1..626
            /organism="Canis familiaris"
            /mol_type="genomic DNA"
            /strain="Standard Poodle"
            /db_xref="taxon:9615"
            /clone_lib="Dog Library"
            /notes="Site 1: BstXI; Libraries were prepared from
            peripheral blood"

FEATURES
    source
        1..626
            /organism="Canis familiaris"
            /mol_type="genomic DNA"
            /strain="Standard Poodle"
            /db_xref="taxon:9615"
            /clone_lib="Dog Library"
            /notes="Site 1: BstXI; Libraries were prepared from
            peripheral blood"

ORIGIN
Query Match      65.1%; Score 21.6; DB 12; Length 626;
Best Local Similarity 56.8%; Pred. No. 3.4e+02;
Matches 21; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY 6 YGNTGGCCNTCCNTAYAARGAYTAYGARGTNAAR 42
Db 549 CACTCATGTCATTCCTCTACAAAGACTATGAA 581

Email: sgn-feedback@sgn.cornell.edu
Insert length: 22763 Std Error: 0.00
Plate: 211 row: C column: 12
Seq primer: SP6
Class: BAC ends
High quality sequence start: 27
High quality sequence stop: 822.

FEATURES
    source
        1..936
            /organism="Lycopersicon esculentum"
            /mol_type="genomic DNA"
            /cultivar="Heinz 1706"
            /db_xref="taxon:4081"
            /clone="LE_HBA0211C12"
            /lab_host="E. coli"
            /clone_lib="Tomato HindIII BAC Library"
            /notes="Vector: pBelOBAC11, Site_1: HindIII"

ORIGIN
Query Match      65.7%; Score 21.8; DB 14; Length 936;
Best Local Similarity 60.6%; Pred. No. 3e+02;
Matches 20; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

QY 4 CAYGNTGGCCNTCCNTAYAARGAYTAYGAR 36
Db 549 CACTCATGTCATTCCTCTACAAAGACTATGAA 581

RESULT 57
DX143003 661 bp DNA linear GSS 18-JAN-2006
LOCUS
DEFINITION
    OG_ABA0177K16.f OG_ABA Oryza granulata genomic clone OG_ABA0177K16
    5', genomic survey sequence.
ACCESSION
    DX143003
VERSION
    DX143003.1 GI:85189269
KEYWORDS
    GSS:
SOURCE
    Oryza granulata
    ORGANISM
        Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
        Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP
        clade; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE
    1 (bases 1 to 661)
    Kim,H., Collura,K., Wissotski,M., Byrne,M., Stum,D., Smart,D.,
    Rao,K., Luo,M., Jetty,R., Kudrna,D., Muller,C., Soderlund,C. and
    Wing,R.
    OMAP (Oryza Map Alignment Project) - Arizona Genomics Institute
    Unpublished (2005)
    JOURNAL
    COMMENT
        Contact: Rod A. Wing
        Email: rwing@genome.arizona.edu
        Tel: 520 621 1259
        Fax: 520 626 9595
        University of Arizona
        Arizona Genomics Institute
        Contact: Rod A. Wing
        Forbes Building Room 303, Tucson, AZ 85721-0036, USA
        Tel: 520 621 1259
        Email: rwing@genome.arizona.edu
        Plate: 0177 row: K column: 16
        Class: BAC ends.
        Location/Qualifiers
            1..661
                /organism="Oryza granulata"
                /mol_type="genomic DNA"
                /db_xref="taxon:110450"
                /clone="OG_ABA0177K16"
                /tissue_type="young leaves"
                /lab_host="DH10B T1 phage resistant"
                /clone_lib="OG_ABA"
                /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"

FEATURES
    source
        1..661
            /organism="Oryza granulata"
            /mol_type="genomic DNA"
            /db_xref="taxon:110450"
            /clone="OG_ABA0177K16"
            /tissue_type="young leaves"
            /lab_host="DH10B T1 phage resistant"
            /clone_lib="OG_ABA"
            /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"

ORIGIN
Query Match      65.1%; Score 21.6; DB 14; Length 661;
Best Local Similarity 56.8%; Pred. No. 3.4e+02;
Matches 21; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY 2 CMCVCGTCGGCCNTCCNTAYAARGAYTAYGARGCT 38
Db 363 CTCTGTCATGGCCATCACCCTATGATCATCAAGT 399

RESULT 58
DX134811 734 bp DNA linear GSS 18-JAN-2006
LOCUS
DEFINITION
    OG_ABA0166J17.f OG_ABA Oryza granulata genomic clone OG_ABA0166J17
    5', genomic survey sequence.
ACCESSION
    DX134811
VERSION
    DX134811.1 GI:85168723
KEYWORDS
    GSS:
SOURCE
    Oryza granulata
    ORGANISM
        Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
        Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP
        clade; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE
    1 (bases 1 to 734)
    Kim,H., Collura,K., Wissotski,M., Byrne,M., Stum,D., Smart,D.,
    Rao,K., Luo,M., Jetty,R., Kudrna,D., Muller,C., Soderlund,C. and
    Wing,R.
    OMAP (Oryza Map Alignment Project) - Arizona Genomics Institute
    Unpublished (2005)
    JOURNAL
    COMMENT
        Contact: Rod A. Wing
        Email: rwing@genome.arizona.edu
        Tel: 520 621 1259
        Fax: 520 626 9595
        University of Arizona
        Arizona Genomics Institute
        Contact: Rod A. Wing
        Forbes Building Room 303, Tucson, AZ 85721-0036, USA
        Tel: 520 621 1259
        Email: rwing@genome.arizona.edu
        Plate: 0177 row: K column: 16
        Class: BAC ends.
        Location/Qualifiers
            1..734
                /organism="Oryza granulata"
                /mol_type="genomic DNA"
                /db_xref="taxon:110450"
                /clone="OG_ABA0166J17"
                /tissue_type="young leaves"
                /lab_host="DH10B T1 phage resistant"
                /clone_lib="OG_ABA"
                /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"

FEATURES
    source
        1..734
            /organism="Oryza granulata"
            /mol_type="genomic DNA"
            /db_xref="taxon:110450"
            /clone="OG_ABA0166J17"
            /tissue_type="young leaves"
            /lab_host="DH10B T1 phage resistant"
            /clone_lib="OG_ABA"
            /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"

ORIGIN
Query Match      65.1%; Score 21.6; DB 14; Length 734;
Best Local Similarity 56.8%; Pred. No. 3.4e+02;
Matches 21; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY 2 CMCVCGTCGGCCNTCCNTAYAARGAYTAYGARGCT 38
Db 363 CTCTGTCATGGCCATCACCCTATGATCATCAAGT 399

```

University of Arizona  
 Forbes Building Room 303, Tucson, AZ 85721-0036, USA  
 Tel: 520 626 9595  
 Fax: 520 621 1259  
 Email: rwing@genome.arizona.edu  
 Plate: 0166 row: J column: 17  
 Class: BAC ends.

## FEATURES

source

Location/Qualifiers  
 1. 734  
 /organism="Oryza granulata"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:110450"  
 /clone="OG ABa0166J17"  
 /tissue types="young leaves"  
 /lab host="DH10B T1 phage resistant"  
 /clone lib="OG\_ABa"  
 /note="Vector: pGIBAC1; Site\_1: HindIII; Site\_2: HindIII"

## ORIGIN

Query Match 65.1%; Score 21.6; DB 14; Length 734;  
 Best Local Similarity 56.8%; Pred. No. 3.5e+02;  
 Matches 21; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

Qy 2 CNYAGNTGGCCNTCNCNTAYAARGATYAGT 38

Db 365 CTCTGTCATGCCCATCACCAATATGTCATCAAGT 401  
 |||||  
 |||||

## RESULT 59

CZ983576

LOCUS

DEFINITION 197310 Tomato MboI BAC Library Lycopersicon esculentum genomic  
 clone SL\_MboI0086A08 5, genomic survey sequence.

ACCESSION

CZ983576

VERSION

CZ983576.1

KEYWORDS

GSS.

SOURCE

Lycopersicon esculentum (Solanum lycopersicum)

ORGANISM

Eukaryota; Viridiplantae;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;

asterids; lamiales; Solanales; Solanaceae; Solanum; Lycopersicon.

1 (bases 1 to 859)

Muller, L.A., Buel, R.M., Wang, Y., Tanksley, S.D., Giovannoni, J.J.,

Van Eck, J., and Stack, S.

BAC end sequencing from three Solanum lycopersicon libraries

Unpublished (2005)

Other\_GSSs: 197309

Contact: Lukas Mueller

Tanksley Lab, Dept. of Plant Breeding

Cornell University

251 Emerson Hall, Ithaca, NY 14853, USA

Tel: 607-255-6557

Fax: 607-255-6683

Email: sgn-feedback@sgn.cornell.edu

Plate: 86 row: A column: 8

Seq primer: T7

Class: BAC ends

High quality sequence start: 59

High quality sequence stop: 316.

Location/Qualifiers

1. 859

/organism="Lycopersicon esculentum"

/mol\_type="genomic DNA"

/cultivar="Heinz 1706"

/db\_xref="taxon:4081"

/clone="SL\_MboI0086A08"

/lab host="E. coli"

/clone lib="Tomato MboI BAC Library"

/note="Vector: pBel0BAC11; Site\_1: MboI"

## ORIGIN

Query Match 65.1%; Score 21.6; DB 13; Length 859;

Best Local Similarity 61.8%; Pred. No. 3.6e+02;

Matches

21; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

Qy 5

AYGNTGGCCNTCNCNTAYAARGATYAGT 38

Db 354

AGCATGGCCACCCCATCCAAAGATTCTGAAGT 387  
 |||||  
 |||||

## RESULT 60

DR753481/c

LOCUS

DEFINITION CCRAP4A04 Coprinus cinereus rapamycin-treated mycelia cDNAs  
 Coprinopsis cinerea cDNA, mRNA sequence.

ACCESSION

DR753481

VERSION

DR753481.1

KEYWORDS

EST.

SOURCE

Coprinopsis cinerea (Coprinus cinereus)

ORGANISM

Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;

Agaricales; Psathyrellaceae; Coprinopsis.

1 (bases 1 to 978)

Carlson, M.D., Barr, C., Murphy, B., Gathman, A.C., Lilly, W.W. and

Pukkila, P.J.

Expressed sequence tags from Coprinus cinereus (Coprinopsis

cinerea) cDNAs, summer 2004

Unpublished (2004)

Contact: Gathman AC

Biology Department

Southeast Missouri State University

1 University Plaza, Cape Girardeau, MO 63701, USA

Tel: 5736512361

Fax: 573 651 2382

Email: agathman@semo.edu.

Location/Qualifiers

1. 978

/organism="Coprinopsis cinerea"

/mol\_type="mRNA"

/strain="Okayama7#130"

/db\_xref="taxon:5346"

/dev stage="vegetative monokaryotic mycelium"

/lab\_host="E. coli XL10-Gold"

/clone lib="Coprinus cinereus rapamycin-treated mycelia

cDNAs"

/notes="Vector: pBluescript II SK+; Site\_1: EcoRI; Site\_2:

XhoI"

## ORIGIN

Query Match 65.1%; Score 21.6; DB 10; Length 978;

Best Local Similarity 53.8%; Pred. No. 3.7e+02;

Matches 21; Conservative 7; Mismatches 11; Indels 0; Gaps 0;

Qy 4

CAYGNTGGCCNTCNCNTAYAARGATYAGTNAAR 42

Db 371

CAGCTTGGCCCTTCCTCTTACAAGGCCCATGGAGTTGAA 333  
 |||||  
 |||||

## RESULT 61

CK084463/c

LOCUS

DEFINITION 79195rsicen.2839.y1 Oryza sativa cv. LYP9 booting whole plant cDNA  
 library Oryza sativa (indica cultivar-group) cDNA 5', mRNA

ACCESSION

CK084463

VERSION

CK084463.1

KEYWORDS

EST.

SOURCE

Oryza sativa (indica cultivar-group)

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP

clade; Ehrhartoideae; Oryzaceae; Oryza.

1 (bases 1 to 583)

Yu, J., Wang, J., Lin, W., Li, S., Li, H., Zhou, J., Ni, P., Dong, W.,

Hu, S., Zeng, C., Zhang, J., Zhang, Y., Li, R., Xu, Z., Li, S., Li, X.,

Zheng, H., Cong, L., Lin, L., Yin, J., Geng, J., Li, G., Shi, J., Liu, J.,

Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,  
Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,  
Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,  
Wu,S. and Liu,J.

The Genomes of *Oryza sativa*: A History of Duplications

PLoS Biol. 3 (2), e38 (2005)

## TITLE

## JOURNAL

## PUBMED

## COMMENT

Contact: Yan Zhou  
Bioinformatics Department  
Hangzhou Genomics Institute  
No.51 ZhiJiang Road, Hangzhou 310008, China  
Tel: 86-571-58805886  
Fax: 86-571-56805884  
Email: zhouyan@genomics.org.cn  
Seq primer: M13 Forward  
High quality sequence stop: 583  
POLYA=No.

## FEATURES

source

Location/Qualifiers

1..583  
/organism="Oryza sativa (indica cultivar-group)"  
/mol\_type="mRNA"  
/cultivar="LYP9"  
/db\_xref="taxon:39946"  
/tissue\_type="whole plant"  
/dev\_stage="booting"  
/clone\_lib="Oryza sativa cv. LYP9 booting whole plant cDNA library"

## ORIGIN

Query Match 64.5%; Score 21.4; DB 5; Length 583;  
Best Local Similarity 53.1%; Pred. No. 4.1e+02;  
Matches 22; Conservative 6; Mismatches 13; Indels 0; Gaps 0;

QY 2 CMCAYCCTGGCCTCCNTCCNTAARGAYTAGTGTAAR 42

Db 100 CGCTCGCCTCGCGCTGCCATCTAAATATTCAGAGGTTTCAG 60

## RESULT 62

## BZ990137

## LOCUS

DEFINITION BZ990137 756 bp DNA linear GSS 25-MAR-2003  
PUGIF847B ZM\_0.6\_1.0\_KB Zea mays genomic clone ZMMBta390M24,  
genomic survey sequence.

## ACCESSION

## BZ990137

## VERSION

## BZ990137.1

## GI:29230635

## GSS.

## SOURCE

## Zea mays

## ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.

## REFERENCE

## AUTHORS

Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,  
Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and  
Bennetzen,J.

## TITLE

## Maize Genomics Consortium

## JOURNAL

## Unpublished (2003)

## COMMENT

## Other GSSs: PUGIF847D

## Contact: Cathy Whitelaw

## TIGR

## 9712 Medical Center Drive, Rockville, MD 20850, USA

## Tel: 301-838-5843

## Fax: 301-838-0208

## Email: whitelaw@tigr.org

## Seq primer: TR

## Class: sheared ends.

## Location/Qualifiers

## 1..756

## /organism="Zea mays"

## /mol\_type="genomic DNA"

## /strain="B73"

## /db\_xref="taxon:4577"

## /clone="ZMMBta390M24"

/clone\_lib="ZM\_0.6\_1.0\_KB"  
/note=Vector: pCR4-TOPO; Site\_1: EcoRI; 0.6-1.0 kb high  
CoT selected genomic DNA library"

## ORIGIN

Query Match 64.5%; Score 21.4; DB 12; Length 756;  
Best Local Similarity 61.3%; Pred. No. 4.3e+02;  
Matches 19; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 12 GCCTCNCNTAYAARGAYTAGTGTAAR 42

Db 645 GCCTTCTCCATATATAAATTATGAGTTAA 675

## RESULT 63

## CC716597/c

## LOCUS

DEFINITION CC716597 774 bp DNA linear GSS 19-JUN-2003  
OGUIZ72TH ZM\_0.7\_1.5\_KB Zea mays genomic clone ZMMBma0452K24,  
genomic survey sequence.

## ACCESSION

## CC716597

## VERSION

## CC716597.1

## GI:32121373

## GSS.

## SOURCE

## Zea mays

## ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.

## REFERENCE

## AUTHORS

Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,  
Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,  
Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.

## TITLE

## Consortium for Maize Genomics

## JOURNAL

## Unpublished (2002)

## COMMENT

## Other GSSs: OGUIZ72TV

## Contact: Cathy Whitelaw

## TIGR

## 9712 Medical Center Drive, Rockville, MD 20850, USA

## Tel: 301-838-5843

## Fax: 301-838-0208

## Email: whitelaw@tigr.org

## Seq primer: TR

## Class: methylation filtered.

## Location/Qualifiers

## 1..774

## /organism="Zea mays"

## /mol\_type="genomic DNA"

## /strain="B73"

## /db\_xref="taxon:4577"

## /clone\_lib="ZMMBma0452K24"

/note=Vector: pBCSK; Site\_1: HincII; 0.7-1.5 kb  
methylation filtered genomic DNA library"

## ORIGIN

Query Match 64.5%; Score 21.4; DB 12; Length 774;  
Best Local Similarity 61.3%; Pred. No. 4.3e+02;  
Matches 19; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 12 GCCTCNCNTAYAARGAYTAGTGTAAR 42

Db 265 GCCTTCTCCATATATAAATTATGAGTTAA 235

## RESULT 64

## CB673669/c

## LOCUS

DEFINITION CB673669 792 bp mRNA linear EST 09-APR-2003  
OSJNEe08E24.f OSJNEe Oryza sativa (japonica cultivar-group) cDNA  
clone OSJNEe08E24 5', mRNA sequence.

## ACCESSION

## CB673669

## VERSION

## CB673669.1

## GI:129677394

## KEYWORDS

## EST.

## SOURCE

## Oryza sativa (japonica cultivar-group)

## ORGANISM

## Oryza sativa (japonica cultivar-group)



Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP clade; Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE  
AUTHORS

1 (bases 1 to 792)

Jantasuriyarat,C., Gowda,M., Haller,K., Hatfield,J., Lu,G., Stahlberg,E., Zhou,B., Li,H., Kim,H., Yu,Y., Dean,R.A., Wing,R.A., Soderlund,C. and Wang,G.L.

Large-scale identification of expressed sequence tags involved in rice and rice blast fungus interaction

Plant Physiol. 138 (1), 105-115 (2005)

JOURNAL  
PUBMED

15888683

## COMMENT

Contact: Rod Wing

Arizona Genomics Institute

University of Arizona

Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ

85721-0088, USA

Tel: 520 626 3967

Fax: 520 621 9288

Email: rwing@genome.arizona.edu

PCR Primers

FORWARD: gta aaa cga cgg cca gtcg

BACKWARD: gga aac agc tat gac cat g

Plate: 08 row: E column: 24

Seq primer: gta aaa cga cgg cca gtcg.

Location/Qualifiers

FEATURES  
source

1..792 /organism="Oryza sativa (japonica cultivar-group)"

/mol\_type="mRNA"

/cultivar="Nipponbare"

/db\_xref="taxon:39947"

/clone="OSJNE08E24"

/tissue\_type="Leaf"

/dev\_stage="3 week"

/lab\_host="DH10B"

/clone\_lib="OSJNE"

/notes="Vector: pBluescript II KS +; Site\_1: EcoRI; Site\_2:

XhoI; 24 hrs after inoculation with Rice Blast (70-15)"

## ORIGIN

Query Match 64.5%; Score 21.4; DB 4; Length 792;

Best Local Similarity 53.7%; Pred. No. 4.3e+02;

Matches 22; Conservative 6; Mismatches 13; Indels 0; Gaps 0;

## RESULT 65

## CB684312/c

## LOCUS

## DEFINITION

OSJNEf13L03.f OSJNEf Oryza sativa (japonica cultivar-group) cDNA

clone OSJNEf13L03 5', mRNA sequence.

## ACCESSION

## CB684312

## VERSION

## CB684312.1

## GI:29688037

## EST.

## SOURCE

## ORGANISM

## Oryza sativa (japonica cultivar-group)

## Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP clade; Ehrhartoideae; Oryzaceae; Oryza.

## 1 (bases 1 to 820)

## Jantasuriyarat,C., Gowda,M., Haller,K., Hatfield,J., Lu,G., Stahlberg,E., Zhou,B., Li,H., Kim,H., Yu,Y., Dean,R.A., Wing,R.A., Soderlund,C. and Wang,G.L.

## Large-scale identification of expressed sequence tags involved in rice and rice blast fungus interaction

## Plant Physiol. 138 (1), 105-115 (2005)

## 15888683

## COMMENT

## Contact: Rod Wing

## Arizona Genomics Institute

## University of Arizona

## Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ

## TITLE

## Large-scale identification of expressed sequence tags involved in rice and rice blast fungus interaction

## Plant Physiol. 138 (1), 105-115 (2005)

## JOURNAL

## PUBMED

## 15888683

## COMMENT

## Contact: Rod Wing

## Arizona Genomics Institute

## University of Arizona

## Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ

85721-0088, USA

Tel: 520 626 3967

Fax: 520 621 9288

Email: rwing@genome.arizona.edu

PCR Primers

FORWARD: gta aaa cga cgg cca gtcg

BACKWARD: gga aac agc tat gac cat g

Plate: 13 row: L column: 03

Seq primer: gta aaa cga cgg cca gtcg.

Location/Qualifiers

FEATURES  
source

1..820

/organism="Oryza sativa (japonica cultivar-group)"

/mol\_type="mRNA"

/cultivar="Nipponbare"

/db\_xref="taxon:39947"

/clone="OSJNEf13L03"

/tissue\_type="Leaf"

/dev\_stage="3 week"

/lab\_host="DH10B"

/clone\_lib="OSJNEf"

/note="Vector: pBluescript II KS +; Site\_1: EcoRI; Site\_2:

XhoI; Uninfected Control"

## ORIGIN

Query Match 64.5%; Score 21.4; DB 4; Length 820;

Best Local Similarity 53.7%; Pred. No. 4.4e+02;

Matches 22; Conservative 6; Mismatches 13; Indels 0; Gaps 0;

QY 2 CNCAYCGTGGCCNTCCNTAYAAAGAYTAYGAGTNAAR 42

Db 622 CGCTCGCTCGCGCCATCTAAATATTCAGAGGTTTAG 582

RESULT 66

CB649010/c

LOCUS

DEFINITION

OSJNEb12L08.f OSJNEb Oryza sativa (japonica cultivar-group) cDNA

clone OSJNEb12L08 5', mRNA sequence.

ACCESSION

CB649010.1

GI:29644003

EST.

KEYWORDS

SOURCE

ORGANISM

Oryza sativa (japonica cultivar-group)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP clade; Ehrhartoideae; Oryzaceae; Oryza.

1 (bases 1 to 821)

Jantasuriyarat,C., Gowda,M., Haller,K., Hatfield,J., Lu,G., Stahlberg,E., Zhou,B., Li,H., Kim,H., Yu,Y., Dean,R.A., Wing,R.A., Soderlund,C. and Wang,G.L.

Large-scale identification of expressed sequence tags involved in rice and rice blast fungus interaction

Plant Physiol. 138 (1), 105-115 (2005)

15888683

Contact: Rod Wing

Arizona Genomics Institute

University of Arizona

Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ

85721-0088, USA

Tel: 520 626 3967

Fax: 520 621 9288

Email: rwing@genome.arizona.edu

PCR Primers

FORWARD: gta aaa cga cgg cca gtcg

BACKWARD: gga aac agc tat gac cat g

Plate: 12 row: L column: 08

Seq primer: gta aaa cga cgg cca gtcg.

Location/Qualifiers

1..821

/organism="Oryza sativa (japonica cultivar-group)"

/mol\_type="mRNA"

/cultivar="Nipponbare"

/db\_xref="taxon:39947"



nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 cDNA library was prepared and sequenced in Mouse Genome  
 Encyclopedia Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
 Division of Experimental Animal Research in Riken contributed to  
 prepare mouse tissues.  
 Tissues were provided by Vassilis Aidinis ( Biomedical Sciences  
 Research Center 'Al. Fleming' Institute of Immunology 14-16 Al.  
 Fleming street 16672 Vari,Greece ) whose assistance we gratefully  
 acknowledge.  
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for  
 further details.

#### FEATURES

source  
 Location/Qualifiers  
 1..343  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="L03008D05"  
 /cell\_type="synovial fibroblasts"  
 /clone\_lib="RIKEN full-length enriched, synovial  
 fibroblasts"  
 ORIGIN  
 Query Match 63.9%; Score 21.2; DB 4; Length 343;  
 Best Local Similarity 58.8%; Pred. No. 4.5e+02;  
 Matches 20; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 4 CAYGNTGCCNTCNCNTAYARGAYTAYGARG 37

DB 121 CACACCTGGCCATCCAGTACAAAACACTCAGG 88

#### RESULT 69

BI861966/c  
 LOCUS  
 DEFINITION 603391623F1 NIH\_MGC\_87 Homo sapiens cDNA clone IMAGE:5406893 5',  
 mRNA sequence.

ACCESSION BI861966.1 GI:16002713

VERSION BI861966.1

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homidae; Homo.

REFERENCE 1 (bases 1 to 615)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: [cgabbs@mail.nih.gov](mailto:cgabbs@mail.nih.gov)

Tissue procurement: DCTD/DTF

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: L1AM12037 row: m column: 06

High quality sequence stop: 591.

#### FEATURES

source  
 Location/Qualifiers  
 1..615  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5406893"  
 /tissue\_type="mammary adenocarcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_87"  
 /notes="Organ: breast; Vector: pCMV-SPORT6; Site 1: NotI;  
 Site 2: SalI; Cloned unidirectionally; oligo-dT primed.  
 Average insert size 1.383 kb. Library enriched for  
 full-length clones and constructed by Life Technologies.

Note: this is a NIH\_MGC Library."

#### ORIGIN

Query Match 63.9%; Score 21.2; DB 2; Length 615;  
 Best Local Similarity 59.0%; Pred. No. 5.1e+02;  
 Matches 23; Conservative 4; Mismatches 12; Indels 0; Gaps 0;  
 QY 4 CAYGNTGCCNTCNCNTAYARGAYTAYGARGTNAAR 42  
 DB 570 CAGGCTGGCACAGCCTTATTAAGATATGATATCAAG 532

#### RESULT 70

CF585003

LOCUS

DEFINITION 1043 bp mRNA linear EST 24-SEP-2003

IMAGE:6430982 5', mRNA sequence.

ACCESSION CF585003

VERSION CF585003.1 GI:35198265

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 1043)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Daniela S. Gerhard, Ph.D.

Office of Cancer Genomics

National Cancer Institute / NIH

Bldg. 31 Rm10A07 Bethesda, MD 20892

Email: [cgabbs@mail.nih.gov](mailto:cgabbs@mail.nih.gov)

Tissue Procurement: Gerard Gradwohl (PNAS 97 P1607-1611, 2000)

cDNA Library Preparation: Catherine Lee, Endocrine Pancreas

Consortium

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: IRBD11 row: f column: 03

High quality sequence start: 20

High quality sequence stop: 596.

#### FEATURES

source

Location/Qualifiers

1..1043

/organism="Mus musculus"

/mol\_type="mRNA"

/db\_xref="taxon:10090"

/clone="IMAGE:6430982"

/lab\_host="DH10B"

/clone\_lib="NIH\_MGC\_137"

/notes="Organ: pancreas; Vector: pSPORT1; Site 1: SalI;  
 Site 2: NotI; Library consists of a pool of clones  
 rearrayed from the following libraries: Melton normalized  
 mixed mouse pancreas 1 N1-MMS1, Amplified Melton mouse  
 islets 1 M1S1-A, and Kaestner ngn3 wt. Clones rearrayed in  
 the laboratory of K. Kaestner (University of  
 Pennsylvania). Note: this is a NIH\_MGC Library."

#### ORIGIN

Query Match 63.9%; Score 21.2; DB 5; Length 1043;

Best Local Similarity 57.5%; Pred. No. 5.6e+02;

Matches 23; Conservative 4; Mismatches 13; Indels 0; Gaps 0;

QY 1 GCNCYGCNTGCCNTCNCNTAYARGAYTAYGARGTNA 40

DB 970 GCCCACACTGGACGCTCTCTTTATATGAACCGAAGAA 1009

#### RESULT 71

AQ461894/c

```

LOCUS      AQ461894              544 bp      DNA      linear      GSS 23-APR-1999
DEFINITION HS_5055_A2_B07_T7A RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=631 Col=14 Row=C, genomic survey sequence.
ACCESSION  AQ461894
VERSION    AQ461894.1 GI:4634664
KEYWORDS   GSS.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homnidae; Homo.
REFERENCE  1 (bases 1 to 544)
AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
PUBMED    10449764
COMMENT    Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieterdejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 631 row: C column: 14
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 544.
FEATURES   source
            1..544
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
                /clone="plate=631 Col=14 Row=C"
                /sex="male"
                /clone_lib="RPCI-11 Human Male BAC Library"
                /notes="vector: pBACE3.6; Site 1: EcoRI; Site 2: EcoRI;
                Male blood DNA was isolated from one randomly chosen donor
                and partially digested with a combination of EcoRI and
                EcoRI Methylase. Size selected DNA was cloned into the
                pBACE3.6 vector at EcoRI sites"
ORIGIN
Query Match      63.3%; Score 21; DB 11; Length 544;
Best Local Similarity 59.5%; Pred. No. 6.1e+02;
Matches 22; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

Qy      6 YGNTGGCCNTCCNTAYARGAYTAGGTAAR 42
      : ||| ||| ||| ||| ||| ||| ||| |||
Db      133 TGTGTCGCCCTCTCCNTGAAAGACTACGACNTAAG 97

RESULT 72
DX336552
LOCUS      DX336552              573 bp      DNA      linear      GSS 20-JAN-2006
DEFINITION OR_Aba0262A04.r OR_Aba Oryza ridleyi genomic clone OR_Aba0262A04
            3', genomic survey sequence.
ACCESSION  DX336552
VERSION    DX336552.1 GI:85613338
KEYWORDS   GSS.
SOURCE     Oryza ridleyi
ORGANISM   Oryza ridleyi
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP
            clade; Ehrhartoideae; Oryzeae; Oryza.

```

```

REFERENCE  1 (bases 1 to 573)
AUTHORS   Kim,H., Collura,K., Wissotski,M., Byrne,M., Stum,D., Smart,D.,
            Rao,K., Luo,M., Jetty,R., Kudrna,D., Muller,C., Soderlund,C. and
            Wing,R.
TITLE     OMAP (Oryza Map Alignment Project)- Arizona Genomics Institute
            Unpublished (2005)
JOURNAL
COMMENT    Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: rwing@genome.arizona.edu
            Plate: 0262 row: A column: 04
            Class: BAC ends.
FEATURES   Location/Qualifiers
            1..573
                /organism="Oryza ridleyi"
                /mol_type="genomic DNA"
                /db_xref="taxon:83308"
                /clone="OR_Aba0262A04"
                /tissue_type="leaves"
                /lab_host="DH10B"
                /clone_lib="OR_Aba"
                /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN
Query Match      63.3%; Score 21; DB 14; Length 573;
Best Local Similarity 61.8%; Pred. No. 6.1e+02;
Matches 21; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

Qy      2 CMCAYCCTGGCCNTCCNTAYARGAYTAGGA 35
      ||| ||| ||| ||| ||| ||| ||| |||
Db      377 CACACATGCACATCTCCATATAGGAGTATGA 410

RESULT 73
AQ727367/c
LOCUS      AQ727367              583 bp      DNA      linear      GSS 14-JUL-1999
DEFINITION HS_5454_A1_D08 SP6E RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=1030 Col=15 Row=G, genomic survey sequence.
ACCESSION  AQ727367
VERSION    AQ727367.1 GI:5487036
KEYWORDS   GSS.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homnidae; Homo.
REFERENCE  1 (bases 1 to 583)
AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
PUBMED    10449764
COMMENT    Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieterdejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 1030 row: G column: 15
            Seq primer: SP6
            Class: BAC ends

```

High quality sequence stop: 583.  
 Location/Qualifiers  
 1..583  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /clone="Plate:1030 Col=15 Row=G"  
 /sex="male"  
 /clone.lib="RPC1-11 Human Male BAC Library"  
 /note="Vector: pBAC3.6; Site 1: EcoRI; Site 2: EcoRI;  
 Male blood DNA was isolated from one randomly chosen donor  
 and partially digested with a combination of EcoRI and  
 EcoRI Methylase. Size selected DNA was cloned into the  
 pBAC3.6 vector at EcoRI sites"

ORIGIN  
 Query Match 63.3%; Score 21; DB 11; Length 583;  
 Best Local Similarity 51.2%; Pred. No. 6.1e+02;  
 Matches 21; Conservative 7; Mismatches 13; Indels 0; Gaps 0;  
 QY 2 CNCAYGCTGGCCNTCCNTAATGAAGATATGATGATGAGTGA 42  
 Db 440 CGATGACGCGCTCCCAAGATATGATGATGAGTGA 400

RESULT 74  
 CW064489  
 LOCUS  
 DEFINITION 602 bp DNA linear GSS 28-OCT-2004  
 104310\_10522599 114\_30141 Sorghum methylation filtered library  
 (LibID: 104) Sorghum bicolor genomic clone 10522599, genomic survey  
 sequence.  
 CW064489  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Sorghum bicolor (sorghum)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Sorghum.  
 1 (bases 1 to 602)  
 Bedell, J.A., Budiman, M.A., Nunberg, A., Citek, R.W., Robbins, D.,  
 Jones, J., Flick, E., Rohlfing, T., Fries, J., Bradford, K.,  
 McMenamy, J., Smith, M., Holman, H., Roe, B.A., Wiley, G., Korf, I.F.,  
 Rabinowicz, P.D., Lakey, N., McCombie, W.R., Jeddloh, J.A. and  
 Martienssen, R.A.  
 Sorghum genome sequencing by methylation filtration  
 PLoS Biol. 3 (1), e13 (2005)  
 15660154  
 Contact: Bedell JA  
 Orion Genomics, LLC  
 4041 Forest Park Ave, St. Louis, MO 63108, USA  
 Tel: 314 615 6379  
 Fax: 314 615 5975  
 Email: jbedell@oriongenomics.com  
 Plate: 310 row: m column: 13  
 Seq primer: M13/pUC Forward  
 Class: methylation filtered  
 High quality sequence stop: 602.  
 Location/Qualifiers  
 1..602  
 /organism="Sorghum bicolor"  
 /mol\_type="genomic DNA"  
 /cultivar="ATx623"  
 /db\_xref="taxon:4558"  
 /clone="10522599"  
 /clone.lib="Sorghum methylation filtered library (LibID:  
 104)"  
 /note="Organ: leaf; Vector: pBCSK(-); Site 1: HincII; DNA  
 prepared from purified nuclei was randomly sheared,  
 end-repaired, size fractionated to enrich for the 0.5 to 5  
 kb fraction, ligated into HincII-digested pBCSK(-) vector  
 and electroporated into E. coli cells. This is a  
 methylation filtered library."

FEATURES  
 source  
 1..602  
 /organism="Sorghum bicolor"  
 /mol\_type="genomic DNA"  
 /cultivar="ATx623"  
 /db\_xref="taxon:4558"  
 /clone="10522599"  
 /clone.lib="Sorghum methylation filtered library (LibID:  
 104)"  
 /note="Organ: leaf; Vector: pBCSK(-); Site 1: HincII; DNA  
 prepared from purified nuclei was randomly sheared,  
 end-repaired, size fractionated to enrich for the 0.5 to 5  
 kb fraction, ligated into HincII-digested pBCSK(-) vector  
 and electroporated into E. coli cells. This is a  
 methylation filtered library."

ORIGIN  
 Query Match 63.3%; Score 21; DB 13; Length 602;  
 Best Local Similarity 58.3%; Pred. No. 6.2e+02;  
 Matches 21; Conservative 5; Mismatches 10; Indels 0; Gaps 0;  
 QY 2 CNCAYGCTGGCCNTCCNTAATGAAGATATGATGAGTGA 37  
 Db 277 CGCCGGGTAGCGCTCGCCGACAGGACGACGACG 312

RESULT 75  
 CN249531  
 LOCUS  
 DEFINITION 615 bp mRNA linear EST 09-APR-2004  
 EST015448 Mycelium and yeast cells from Paracoccidioides  
 brasiliensis Paracoccidioides brasiliensis cDNA, mRNA sequence.  
 CN249531  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Paracoccidioides brasiliensis  
 Paracoccidioides brasiliensis  
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 Onygenales; Mitosporic Onygenales; Paracoccidioides.  
 1 (bases 1 to 615)  
 Felipe, M.S., Andrade, R.V., Arraes, F.B., Nicola, A.M., Matanhao, A.Q.,  
 Torres, F.A., Silva-Pereira, I., Pocas-Fonseca, M.J., Campos, E.G.,  
 Moraes, L.M., Andrade, P.A., Tavares, A.H., Silva, S.S., Kyaw, C.M.,  
 Souza, D.P., Network, P., Pereira, M., Jesuino, R.S., Andrade, E.V.,  
 Parente, J.A., Oliveira, G.S., Barbosa, M.S., Martins, N.F.,  
 Pachin, A.L., Cardoso, R.S., Passos, G.A., Almeida, N.F., Walter, M.E.,  
 Soares, C.M., Carvalho, M.J. and Brígido, M.M.  
 Transcriptional Profiles of the Human Pathogenic Fungus  
 Paracoccidioides brasiliensis in Mycelium and Yeast Cells  
 J. Biol. Chem. 280 (26), 24706-24714 (2005)  
 15849188  
 Contact: Felipe MSS  
 Laboratory of Molecular Biology  
 Institute of Biology - University of Brasilia  
 Campus Universitario, Asa Norte, Brasilia, DF 70910-900, BRA  
 Tel: 55 61 307 2423  
 Fax: 55 61 349 8411  
 Email: msueli@unb.br  
 Seq primer: T7 Sequencing primer.  
 Location/Qualifiers  
 1..615  
 /organism="Paracoccidioides brasiliensis"  
 /mol\_type="mRNA"  
 /strain="Pb01"  
 /db\_xref="taxon:121759"  
 /clone.lib="Mycelium and yeast cells from Paracoccidioides  
 brasiliensis"  
 /note="Pb Lambda Zap Express Library"

ORIGIN  
 Query Match 63.3%; Score 21; DB 8; Length 615;  
 Best Local Similarity 51.2%; Pred. No. 6.2e+02;  
 Matches 21; Conservative 7; Mismatches 13; Indels 0; Gaps 0;  
 QY 2 CNCAYGCTGGCCNTCCNTAATGAAGATATGATGAGTGA 42  
 Db 492 CTCATTCTCGCCGCTCCCAATATAATCTACGCACTCAA 532

RESULT 76  
 CZ770405/c  
 LOCUS  
 DEFINITION 659 bp DNA linear GSS 26-JUL-2005  
 OC\_Ba0127H15.r OC\_Ba Oryza coarctata genomic clone OC\_Ba0127H15  
 3'\_genomic survey Sequence.  
 CZ770405  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Oryza coarctata (Porteresia coarctata)  
 Oryza coarctata



## ORIGIN

Query Match 63.3%; Score 21; DB 13; Length 687;  
 Best Local Similarity 58.3%; Pred.No. 6.3e+02;  
 Matches 21; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

Qy 2 CNCAYGCGTGGCCNTCCNTAYARGATYAYGARG 37  
 Db 149 CGCGCGGTAGCGCTGCGCCGACAGGACGACGAC 114

## RESULT 79

DX201680 751 bp DNA linear GSS 19-JAN-2006  
 LOCUS OR\_Aba0071B03.f OR\_Aba Oryza ridleyi genomic clone OR\_Aba0071B03  
 DEFINITION 5', genomic survey sequence.

ACCESSION DX201680  
 VERSION DX201680.1 GI:85451106  
 KEYWORDS GSS.

## SOURCE

ORGANISM Oryza ridleyi  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP  
 clade; Ehrhartoideae; Oryzaceae; Oryza.

## REFERENCE

AUTHORS Kim.H., Collura,K., Missotski,M., Byrne,M., Stum,D., Smart,D.,  
 Rao,K., Luo,M., Jetty,R., Kudrna,D., Muller,C., Soderlund,C. and  
 Wing,R.

TITLE ONAP (Oryza Map Alignment Project) - Arizona Genomics Institute  
 JOURNAL Unpublished (2005)

## COMMENT

Contact: Rod A. Wing  
 Arizona Genomics Institute  
 University of Arizona  
 Forbes Building Room 303, Tucson, AZ 85721-0036, USA  
 Tel: 520 626 9595  
 Fax: 520 621 1259  
 Email: rwing@genome.arizona.edu  
 Plate: 0071 row: B column: 03  
 Class: BAC ends.

## FEATURES

source  
 1..751  
 /organism="Oryza ridleyi"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:83308"  
 /clone="OR\_Aba0071B03"  
 /tissue\_type="leaves"  
 /lab\_host="DH10B"  
 /clone\_lib="OR\_ABa"  
 /notes="Vector: pAGIBAC1; Site\_1: HindIII; Site\_2: HindIII"

## ORIGIN

Query Match 63.3%; Score 21; DB 14; Length 751;  
 Best Local Similarity 55.3%; Pred.No. 6.5e+02;  
 Matches 21; Conservative 6; Mismatches 11; Indels 0; Gaps 0;

Qy 4 CAYCGTGGCCNTCCNTAYARGATYAYGARGTNA 41  
 Db 365 CTTGCATGGCCATCACCCCATCCGATTACCAAGTGCA 402

## RESULT 80

DW547742/c 794 bp mRNA linear EST 17-JAN-2006  
 LOCUS EST\_ssal\_rgb2\_12161\_rgb2 Salmo salar cDNA clone  
 DEFINITION ssal\_rgb2\_520\_196\_rev 5', mRNA sequence.

ACCESSION DW547742  
 VERSION DW547742.1 GI:85019086  
 KEYWORDS EST.

## SOURCE

ORGANISM Salmo salar (Atlantic salmon)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
 Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.

## REFERENCE

1 (bases 1 to 794)  
 Koop,B.F., Davidson,W.S. and cGRASP Consortium.  
 Expressed Sequence Tags from Atlantic Salmon Unpublished (2006)  
 JOURNAL Unpublished (2006)  
 COMMENT Contact: Koop BF

Centre for Biomedical Research  
 University of Victoria  
 PO Box 3020 STN CSC, Victoria BC, V8W 3N5, Canada

Tel: 250 472 4067  
 Fax: 250 472 4075  
 Email: bkoop@uvic.ca

Sequenced by: Genome Sciences Centre, BC Cancer Agency Rob Holt, M.  
 Marra. Bioinformatics: Centre for Biomedical Research, University  
 of Victoria Jong Leong, BF Koop.

Insert Length: 794 Std Error: 0.00

Plate: 520

Seq primer: M13 Reverse

High quality sequence stop: 794.

## FEATURES

source  
 1..794  
 /organism="Salmo salar"  
 /mol\_type="mRNA"  
 /strain="McConnell"  
 /db\_xref="taxon:8030"  
 /clone="ssal\_rgb2\_520\_196\_rev"  
 /tissue\_type="mixed tissue"  
 /clone\_lib="rgb2"  
 /note="Organ: brain, kidney, spleen; Vector: pCMVaport6;  
 ssalrgb2 mixed tissue Salmo salar cDNA; Tissue  
 contributors: Robert Devlin (DFO, Vancouver, B.C.)"

## ORIGIN

Query Match 63.3%; Score 21; DB 10; Length 794;  
 Best Local Similarity 55.3%; Pred.No. 6.5e+02;  
 Matches 21; Conservative 6; Mismatches 11; Indels 0; Gaps 0;

Qy 1 GCNCAYCGTGGCCNTCCNTAYARGATYAYGARGT 38  
 Db 650 GCACACTTGGCACTCTCCACACAGCCCTCATGAGT 613

## RESULT 81

## LOCUS

CW825776 813 bp DNA linear GSS 01-SEP-2005  
 Ynhw9057 HW-YUBAC Bos taurus genomic clone  
 DEFINITION HW-YUBAC2-107-03-E09-T7.ab1 5', genomic survey sequence.

ACCESSION CW825776

VERSION CW825776.1 GI:74045687

KEYWORDS GSS.

SOURCE Bos taurus (cattle)

## ORGANISM

Bos taurus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;  
 Pecora; Bovidae; Bovinae; Bos.

1 (bases 1 to 813)

Choi,I.

Genomic sequences from Korean Cattle (Hanwoo) blood

Unpublished (2003)

Contact: Inho Choi

Molecular Biology

Yeungnam University

214-1, Dae-dong Gyeongsan, Korea 712-749

Tel: 82 53 810 2933

Fax: 82 53 816 3637

Email: inhochoi@yumail.ac.kr

Seq primer: T7 Forward

Class: BAC ends.

## FEATURES

source  
 1..813  
 /organism="Bos taurus"  
 /mol\_type="genomic DNA"  
 /strain="Korean Cattle (Hanwoo)"  
 /db\_xref="taxon:9913"

```

/clone="HW-YUBAC2-107-03--E09-T7.ab1"
/sex="Male"
/tissue_type="Blood"
/cell_type="Leucocyte"
/dev_stage="Adult"
/clone_lib="HW-YUBAC"
/note="Vector: pIndigoBAC-5"

ORIGIN
Query Match      63.3%; Score 21; DB 13; Length 813;
Best Local Similarity 58.3%; Pred. No. 6.6e+02;
Matches 21; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

QY 6 YGNCGTGCCNTCCNTAYAARGAYTAYGARGTNA 41
    :|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
432 TGCATGCCCATCACCAACCATGATGAATGAATTAA 467

RESULT 82
CZ881566/c
LOCUS
DEFINITION
OC_Ba0284F06.r OC_Ba Oryza coactata genomic clone OC_Ba0284F06
3'', genomic survey sequence.
ACCESSION
CZ881566
VERSION
CZ881566.1 GI:71357322
KEYWORDS
GSS.
SOURCE
Oryza coactata (Porteresia coactata)
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP
clade; Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 843)
Kim H., Collura, K., Wisotski, M., Byrne, M., Stum, D., Smart, D.,
Rao, K., Luo, M., Jetty, R., Kudrna, D., Muller, C., Soderlund, C. and
Wing, R.
OMAP (Oryza Map Alignment Project)- Arizona Genomics Institute
Unpublished (2005)
Contact: Rod A. Wing
Arizona Genomics Institute
University of Arizona
Forbes Building Room 303, Tucson, AZ 85721-0036, USA
Tel: 520 626 9595
Fax: 520 621 1259
Email: rwing@genome.arizona.edu
PCR Primers
FORWARD: TAA TAC GAC TCA CTA TAG GG
BACKWARD: CAC TCA TTA GGC ACC CCA
Plate: 0284 row: F column: 06
Seq primer: CAC TCA TTA GGC ACC CCA
Class: BAC ends.

FEATURES
            Location/Qualifiers
source          1..843
                /organism="Oryza coactata"
                /mol_type="genomic DNA"
                /db_xref="taxon:77588"
                /clone="OC_Ba0284F06"
                /tissue_type="leaves"
                /dev_stage="mature"
                /lab_host="DH10B"
                /clone_lib="OC_Ba"
                /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"

ORIGIN
Query Match      63.3%; Score 21; DB 13; Length 843;
Best Local Similarity 55.3%; Pred. No. 6.6e+02;
Matches 21; Conservative 6; Mismatches 11; Indels 0; Gaps 0;

QY 4 CAYGCGTGGCCNTCCNTAYAARGAYTAYGARGTNA 41
    :|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
296 CATGATGTCATCAACCATGATGAATGAATGAATGA 259

RESULT 83

```



7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan  
 Tel: 81-03-5841-4434  
 Fax: 81-03-5841-4434  
 Email: m\_taira@biol.s.u-tokyo.ac.jp,  
 URL: <http://www.shigen.nig.ac.jp/nbrp/xenopus/est/>.  
 Location/Qualifiers

## FEATURES

source

1..369  
 /organism="Xenopus laevis"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:8355"  
 /clone="X1500ml6x"  
 /tissue\_type="anterior neuroectoderm"  
 /dev\_stage="late gastrula (stage 12.5)"  
 /clone\_lib="Osada Taira anterior neuroectoderm (ANE)  
 pCS105 cDNA library"

## ORIGIN

Query Match 62.7%; Score 20.8; DB 3; Length 369;  
 Best Local Similarity 52.4%; Pred. No. 6.9e+02;  
 Matches 22; Conservative 6; Mismatches 14; Indels 0; Gaps 0;

QY 1 GCNCAYGCGTCCNTCNCNTAYAAAGAYTAYGARGTNAAR 42  
 |||||  
 Db 219 GCTCATGCTTGGCAACCCCTGGAACCAAGAAATGCGAAGTAAAG 260

## RESULT 85

AV802339

LOCUS

DEFINITION AV802339 RAFL9 Arabidopsis thaliana cDNA clone RAFL09-31-123 3',  
 mRNA sequence.

ACCESSION AV802339

VERSION

KEYWORDS

SOURCE

ORGANISM

Arabidopsis thaliana (thale cress)  
 Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

## REFERENCE

AUTHORS

1 (bases 1 to 435)  
 Seki, M., Narusaka, M., Ishida, J., Kamiya, A., Satou, M., Nakajima, M.,  
 Oono, Y., Sakurai, T., Carninci, P., Kawai, J., Itoh, M., Ishii, Y.,  
 Arakawa, T., Shibata, K., Shinagawa, A., Muramatsu, M., Hayashizaki, Y.  
 and Shinozaki, K.

## TITLE

JOURNAL

COMMENT

Large scale analysis of Arabidopsis full-length cDNA (2002b)  
 Unpublished (2002)  
 Contact: Motoaki Seki  
 Plant Functional Genomics Research Group  
 RIKEN Genomic Sciences Center  
 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  
 Tel: 81-298-36-4359  
 Fax: 81-298-36-9060

Email: msekic@rc.riken.go.jp  
 An Arabidopsis full-length cDNA library was constructed essentially  
 as reported previously (Seki et al., 1998). cDNA cleaved with BamHI  
 and XhoI was ligated to modified Lambda FLC-1 vector (Carninci et  
 al., submitted for publication) digested with BamHI and SalI. This  
 site is in a modified pluscript vector. Please visit our web  
 site ([http://www.gsc.riken.go.jp/e/plant/index\\_e.html](http://www.gsc.riken.go.jp/e/plant/index_e.html)) for further  
 details.

## FEATURES

source

1..435  
 /organism="Arabidopsis thaliana"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:3702"  
 /clone="RAFL09-31-123"  
 /dev\_stage="plants at various developmental stages from  
 germination to mature seeds"  
 /lab\_host="DH10B"  
 /clone\_lib="RAFL9"  
 /notes="Site 1: BamHI; Site 2: SalI; subjected to  
 dehydration (1, 2, 5, 10, 24 hr) and cold (1, 2, 5, 10, 24  
 hr) treatments"

## ORIGIN

Query Match 62.7%; Score 20.8; DB 7; Length 435;  
 Best Local Similarity 59.5%; Pred. No. 7.1e+02;  
 Matches 22; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

QY 4 CAYGCGTCCNTCNCNTAYAAAGAYTAYGARGTNA 40  
 |||||  
 Db 320 CATGCATTACCTCCCTCCCAACAAAGAAATGGAAGTCA 356

## RESULT 86

AZ177524

LOCUS

DEFINITION

SP 0147\_A1 H10 SP6E Strongylocentrotus purpuratus, purple sea  
 urchin, sperm genomic BAC library Strongylocentrotus purpuratus  
 genomic clone Plate=147 Col=19 Row=O, genomic survey sequence.

ACCESSION AZ177524

VERSION

KEYWORDS

SOURCE

ORGANISM

Strongylocentrotus purpuratus  
 Strongylocentrotus purpuratus  
 Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
 Echinoidea; Euechinoidea; Echinacea; Echinoida;  
 Strongylocentrotidae; Strongylocentrotus.

## REFERENCE

AUTHORS

1 (bases 1 to 498)  
 Cameron, R.A., Mahairas, G., Rast, J.P., Martinez, P., Biondi, T.R.,  
 Swartzell, S., Wallace, J.C., Pousetka, A.J., Livingston, B.T.,  
 Wray, G.A., Ettensohn, C.A., Lehrach, H., Britten, R.J., Davidson, E.H.  
 and Hood, L.

## TITLE

JOURNAL

PUBMED

COMMENT

A sea urchin genome project: Sequence scan, virtual map, and  
 additional resources  
 Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)  
 10920195  
 Contact: Cameron, RA, Davidson, EH, Hood, L  
 Division of Biology 156-29  
 California Institute of Technology  
 Pasadena California 91125, USA

Tel: (626) 395-8421

Fax: (626) 793-3047

Email: acameron@caltech.edu

Plate: 147 row: O column: 19

Seq primer: SP6

Class: BAC ends

High quality sequence stop: 498.

## FEATURES

source

Location/Qualifiers

1..498  
 /organism="Strongylocentrotus purpuratus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:7668"  
 /clone="plate=147 Col=19 Row=O"  
 /clone\_lib="Strongylocentrotus purpuratus, purple sea  
 urchin, sperm genomic BAC library"  
 /note="Organ: sperm; Vector: BACE3.6; BAC Clones in E-Coli  
 DH10B"

## ORIGIN

Query Match 62.7%; Score 20.8; DB 11; Length 498;  
 Best Local Similarity 54.8%; Pred. No. 7.3e+02;  
 Matches 23; Conservative 6; Mismatches 13; Indels 0; Gaps 0;

QY 1 GCNCAYGCGTCCNTCNCNTAYAAAGAYTAYGARGTNAAR 42

Db 53 GCAGACGCTGCCGTCNACATTTAAGGACACCAATTATAG 94

## RESULT 87

DN752781

LOCUS

DEFINITION

GL-Ct-9576 GLGC-LIB0001-cf Canis familiaris Normalized Mixed tissue  
 cDNA Library Canis familiaris cDNA, mRNA sequence.

ACCESSION DN752781

VERSION

DN752781.1

GI:62141894

```

KEYWORDS
SOURCE  EST.
ORGANISM  Canis familiaris (dog)
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
          Canis.

REFERENCE
AUTHORS  Liu,Q.
TITLE    Direct Submission Gene Logic Inc
JOURNAL  Unpublished (2005)
COMMENT  Contact: Qing Liu
          Gene Logic Inc.
          610 Professional Drive, Gaithersburg, MD 20879, USA
          Tel: 301 987 1700
          Email: qlu@genelegic.com.

FEATURES
source
1. .581
   /organism="Canis familiaris"
   /mol_type="mRNA"
   /strain="beagle"
   /db_xref="taxon:9615"
   /lab_host="EMDH10B"
   /clone_lib="GLGC-LTB0001-cf Canis familiaris Normalized
Mixed Tissue cDNA Library"
   /notes="Organ: heart, liver, kidney, testis, and brain;
Vector: pCMVSPORT6.0; Site_1: NotI; Site_2: SalI"

ORIGIN
Query Match      62.7%; Score 20.8; DB 9; Length 581;
Best Local Similarity 55.0%; Pred. No. 7.5e+02;
Matches 22; Conservative 5; Mismatches 13; Indels 0; Gaps 0;

QY 1 GCNCAYGCNTGCCNTCCNTAYAARGAYTAYGARGTNA 40
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 339 GCCATGCGAGGCCCTCACACACAGGCGCTACGCTCTTA 378

RESULT 88
DR3K23S/c
LOCUS      590 bp DNA linear GSS 27-NOV-2002
DEFINITION Danio rerio genomic clone DKEY-3K23, genomic survey sequence.
ACCESSION  AL740353
VERSION    AL740353.1 GI:213511134
KEYWORDS  GSS.
SOURCE    Danio rerio (zebrafish)
ORGANISM  Danio rerio
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
          Cypriniformes; Cyprinidae; Danio.

REFERENCE
AUTHORS  Humphray,S.J., Huckle,E. and Hunt,S.E.
TITLE    Direct Submission
JOURNAL  Submitted (06-JUN-2002) The Sanger Institute, Wellcome Trust Genome
          Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail contact:
          humquerry@sanger.ac.uk Unpublished
COMMENT  This sequence was generated from the SP6 end of BAC 3K23. 3K23 is
          part of the Daniokey BAC Library created by R. Plasterk and N.V.
          Keygene.
          Further details: http://www.sanger.ac.uk/Projects/D_rerio/.

FEATURES
source
1. .590
   /organism="Danio rerio"
   /mol_type="genomic DNA"
   /db_xref="taxon:7955"
   /clone="DKEY-3K23"
   /tissue_type="testis"
   /notes="vector pindigobAC-536"

ORIGIN
Query Match      62.7%; Score 20.8; DB 14; Length 590;
Best Local Similarity 52.4%; Pred. No. 7.5e+02;
Matches 22; Conservative 6; Mismatches 14; Indels 0; Gaps 0;

KEYWORDS
SOURCE  1 GCNCAYGCNTGCCNTCCNTAYAARGAYTAYGARGTNAAR 42
        ||||| ||||| ||||| ||||| ||||| ||||| |||||
        268 GCGTGTGCAAGCCCTCTTCTTCAAGCGCTATGATGAAA 227

RESULT 89
CD496729
LOCUS      647 bp mRNA linear EST 12-JUN-2003
DEFINITION CDA23-H06.x1d-t SHGC-CDA Gasterosteus aculeatus cDNA clone
          CDA23-H06 5', mRNA sequence.
ACCESSION  CD496729
VERSION    CD496729.1 GI:31423760
KEYWORDS  EST.
SOURCE    Gasterosteus aculeatus (three spined stickleback)
          Gasterosteus aculeatus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
          Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes;
          Gasterosteidae; Gasterosteus.
          1 (bases 1 to 647)
          Kingsley,D.M., Peichel,C., Balabhadra,S., Grimwood,J., Dickson,M.,
          Schmutz,J. and Myers,R.M.
          Expressed sequence tags from Gasterosteus aculeatus
          Unpublished (2003)
          Contact: Kingsley, DM
          HHMI and Department of Developmental Biology
          Stanford University School of Medicine
          Beckman Center B300, 279 Campus Drive, Stanford, CA 94305-5329, USA
          Tel: 650 725 5954
          Fax: 650 725 7739
          Email: kingsley@cngm.stanford.edu
          Plate: 23
          High quality sequence stop: 647.

FEATURES
source
1. .647
   /organism="Gasterosteus aculeatus"
   /mol_type="mRNA"
   /strain="Salinas river, CA"
   /db_xref="taxon:69293"
   /clone="CDA23-H06"
   /sex="mixed male and female"
   /tissue_type="heads and internal organs combined"
   /dev_stage="adult"
   /clone_lib="SHGC-CDA"
   /note="Vector: lambda ZAP Express/pBK-CMV; Site_1: EcoRI
(5' adaptor); Site_2: XhoI (3' linker primer); The mixed
organ cDNA library was generated using the ZAP-cDNA method
by Stratagene. First strand cDNA synthesis was primed with
a 50 bp linker primer containing an oligo dT sequence
preceded by a synthetic XhoI site. 5 prime adaptors were
used containing an EcoRI cohesive end. The finished cDNAs
were inserted in to the ZAP express vector
unidirectionally in the sense orientation with respect to
the lacZ promoter of pBK-CMV. An amplified library was
prepared from approximately 3 million primary clones in
the lambda ZAP Express vector. In vivo excision was then
used to generate individual pBK-CMV phagemid clones for
EST sequencing."

ORIGIN
Query Match      62.7%; Score 20.8; DB 5; Length 647;
Best Local Similarity 52.4%; Pred. No. 7.7e+02;
Matches 22; Conservative 6; Mismatches 14; Indels 0; Gaps 0;

QY 1 GCNCAYGCNTGCCNTCCNTAYAARGAYTAYGARGTNAAR 42
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 95 GCCCATGTTGGATTTCCTCGTACAGAAAATGAATATAA 136

RESULT 90
CW098794/c
LOCUS      669 bp DNA linear GSS 28-OCT-2004
DEFINITION 104_465_11003200_116_34373_056 Sorghum methylation filtered library

```

(LibID: 104) Sorghum bicolor genomic clone 11003200, genomic survey sequence.

ACCESSION CW098794  
VERSION CW098794.1 GI:54775524

KEYWORDS GSS.  
SOURCE Sorghum bicolor (sorghum)

ORGANISM Sorghum bicolor

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Sorghum.

1 (bases 1 to 669)

Bedell, J.A., Budiman, M.A., Nunberg, A., Citek, R.W., Robbins, D., Jones, J., Flick, E., Rohlfing, T., Fries, J., Bradford, K., McMenamy, J., Smith, M., Holuman, H., Roe, B.A., Wiley, G., Korf, I.F., Rabinowicz, P.D., Lakey, N., McCombie, W.R., Jeddeloh, J.A. and Martienssen, R.A.

Sorghum genome sequencing by methylation filtration

PLoS Biol. 3 (1), e13 (2005)

15660154

COMMENT Contact: Bedell JA

Orion Genomics, LLC

4041 Forest Park Ave, St. Louis, MO 63108, USA

Tel: 314 615 6979

Fax: 314 615 5975

Email: jbedell@oriongenomics.com

Plate: 465 row: i column: 16

Seq primer: T3 Reverse

Class: methylation filtered

High quality sequence stop: 669.

FEATURES  
source

1. .669

/organism="Sorghum bicolor"

/mol\_type="genomic DNA"

/cultivar="ATx623"

/db\_xref="taxon:4558"

/clone\_lib="11003200"

/clone\_lib="Sorghum methylation filtered library (LibID: 104)"

/notes="Organ: leaf; Vector: pBCSK(-); Site 1: HincII; DNA prepared from purified nuclei was randomly sheared, end-repaired, size fractionated to enrich for the 0.5 to 5 kb fraction, ligated into HincII-digested pBCSK(-) vector and electroporated into E. coli cells. This is a methylation filtered library."

ORIGIN

Query Match 62.7%; Score 20.8; DB 13; Length 669;

Best Local Similarity 52.4%; Pred. No. 7.7e+02;

Matches 22; Conservative 6; Mismatches 14; Indels 0; Gaps 0;

QY 1 GCNCAYGCNTGGCCNTCNCNTAYAGAYTAYGARGTNAAR 42

Db 416 GCCCGCGCTGGGCTTCGCGGTTCAACGATTCACAGAGAA 375

RESULT 91

CW09783

LOCUS

DEFINITION 104\_798\_11469373\_148\_36240\_051 Sorghum methylation filtered library (LibID: 104) Sorghum bicolor genomic clone 11469373, genomic survey sequence.

ACCESSION CW09783

VERSION CW09783.1 GI:55025971

KEYWORDS GSS.

SOURCE Sorghum bicolor

ORGANISM Sorghum bicolor

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Sorghum.

1 (bases 1 to 677)

Bedell, J.A., Budiman, M.A., Nunberg, A., Citek, R.W., Robbins, D., Jones, J., Flick, E., Rohlfing, T., Fries, J., Bradford, K., McMenamy, J., Smith, M., Holuman, H., Roe, B.A., Wiley, G., Korf, I.F.,

Rabinowicz, P.D., Lakey, N., McCombie, W.R., Jeddeloh, J.A. and Martienssen, R.A.

Sorghum genome sequencing by methylation filtration

PLoS Biol. 3 (1), e13 (2005)

15660154

COMMENT Contact: Bedell JA

Orion Genomics, LLC

4041 Forest Park Ave, St. Louis, MO 63108, USA

Tel: 314 615 6979

Fax: 314 615 5975

Email: jbedell@oriongenomics.com

Plate: 798 row: m column: 13

Seq primer: SWfor Forward

Class: methylation filtered

High quality sequence stop: 677.

FEATURES  
source

1. .677

/organism="Sorghum bicolor"

/mol\_type="genomic DNA"

/cultivar="ATx623"

/db\_xref="taxon:4558"

/clone\_lib="11469373"

/clone\_lib="Sorghum methylation filtered library (LibID: 104)"

/notes="Organ: leaf; Vector: pBCSK(-); Site 1: HincII; DNA prepared from purified nuclei was randomly sheared, end-repaired, size fractionated to enrich for the 0.5 to 5 kb fraction, ligated into HincII-digested pBCSK(-) vector and electroporated into E. coli cells. This is a methylation filtered library."

ORIGIN

Query Match 62.7%; Score 20.8; DB 13; Length 677;

Best Local Similarity 52.4%; Pred. No. 7.8e+02;

Matches 22; Conservative 6; Mismatches 14; Indels 0; Gaps 0;

QY 1 GCNCAYGCNTGGCCNTCNCNTAYAGAYTAYGARGTNAAR 42

Db 546 GCCCGCGCTGGGCTTCGCGGTTCAACGATTCACAGAGAA 587

RESULT 92

CW09782/c

LOCUS

DEFINITION 104\_798\_11469373\_116\_36239\_051 Sorghum methylation filtered library (LibID: 104) Sorghum bicolor genomic clone 11469373, genomic survey sequence.

ACCESSION CW09782

VERSION CW09782.1 GI:55025970

KEYWORDS GSS.

SOURCE Sorghum bicolor

ORGANISM Sorghum bicolor

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Sorghum.

1 (bases 1 to 685)

Bedell, J.A., Budiman, M.A., Nunberg, A., Citek, R.W., Robbins, D.,

Jones, J., Flick, E., Rohlfing, T., Fries, J., Bradford, K.,

McMenamy, J., Smith, M., Holuman, H., Roe, B.A., Wiley, G., Korf, I.F.,

Rabinowicz, P.D., Lakey, N., McCombie, W.R., Jeddeloh, J.A. and

Martienssen, R.A.

Sorghum genome sequencing by methylation filtration

PLoS Biol. 3 (1), e13 (2005)

15660154

COMMENT Contact: Bedell JA

Orion Genomics, LLC

4041 Forest Park Ave, St. Louis, MO 63108, USA

Tel: 314 615 6979

Fax: 314 615 5975

Email: jbedell@oriongenomics.com

Plate: 798 row: m column: 13

Seq primer: T3 Reverse

Class: methylation filtered

FEATURES		High quality sequence stop: 685.	
source		Location/Qualifiers	
		1. .685	
		/organism="Sorghum bicolor"	
		/mol_type="genomic DNA"	
		/cultivar="ATx823"	
		/db_xref="taxon:4558"	
		/clone="11469373"	
		/clone_lib="Sorghum methylation filtered library (LibID: 104)"	
		/note="Organ: leaf; Vector: pBCSK(-); Site 1: HincII; DNA prepared from purified nuclei was randomly sheared, end-repaired, size fractionated to enrich for the 0.5 to 5 kb fraction, ligated into HincII-digested pBCSK(-) vector and electroporated into E. coli cells. This is a methylation filtered library."	
ORIGIN			
Query Match		62.7%; Score 20.8; DB 13; Length 685;	
Best Local Similarity		52.4%; Pred. No. 7.8e+02;	
Matches		22; Conservative 6; Mismatches 14; Indels 0; Gaps 0;	
QY		1 GCNCAYGNTGCCNTCNCNTAYAAAGAYTAGGTNAAR 42	
Db		160 GCCCGCGTGGTCTTCGCCGTTCAACGATTCAAAGAGAA 119	
RESULT 93			
CA790172		861 bp mRNA linear EST 04-DEC-2002	
LOCUS		AGENCOURT_10309392 NICHD_XGC_Emb1 Xenopus laevis cDNA clone	
DEFINITION		IMAGE:5156169 5', mRNA sequence.	
ACCESSION		CA790172	
VERSION		CA790172.1 GI:26035894	
KEYWORDS		EST.	
SOURCE		Xenopus laevis (African clawed frog)	
ORGANISM		Xenopus laevis	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoides; Pipidae; Xenopodinae; Xenopus; Xenopus.	
AUTHORS		1 (bases 1 to 861)	
TITLE		NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.	
JOURNAL		National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Unpublished (1997)	
COMMENT		Contact: Robert Strausberg, Ph.D. Email: cgaps-r@mail.nih.gov Tissue Procurement: Martha Rebert, Steven L. Klein, Ph.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM1386 row: n column: 10 High quality sequence stop: 614.	
FEATURES		Location/Qualifiers	
source		1. .861	
		/organism="Xenopus laevis"	
		/mol_type="mRNA"	
		/clone="IMAGE:5156169"	
		/tissue_type="embryo (stage 10)"	
		/lab_host="DH10B (phage-resistant)"	
		/clone_lib="NICHD_XGC_Emb1"	
		/note="Vector: pCMV-SF0RT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.55 kb. Constructed by Life Technologies. Note: This is a Xenopus Gene Collection (XGC) library."	
ORIGIN			
Query Match		62.7%; Score 20.8; DB 4; Length 861;	
Best Local Similarity		52.4%; Pred. No. 8.1e+02;	
Matches		22; Conservative 6; Mismatches 14; Indels 0; Gaps 0;	
QY		1 GCNCAYGNTGCCNTCNCNTAYAAAGAYTAGGTNAAR 42	
Db		160 GCCCGCGTGGTCTTCGCCGTTCAACGATTCAAAGAGAA 119	
RESULT 94			
CA790172		906 bp mRNA linear EST 19-JAN-2006	
LOCUS		CNB362-G10 Y1d-s SHGC-CNB2 Gasterosteus aculeatus cDNA clone	
DEFINITION		CNB362-G10 5', mRNA sequence.	
ACCESSION		DW670971	
VERSION		DW670971.1 GI:85469653	
KEYWORDS		EST.	
SOURCE		Gasterosteus aculeatus (three spined stickleback)	
ORGANISM		Gasterosteus aculeatus	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes; Gasterosteidae; Gasterosteus.	
AUTHORS		1 (bases 1 to 906)	
TITLE		Kingsley, D.M., Peichel, C., Knecht, A., Balabhadra, S., Grimwood, J., Dickson, M., Schmutz, J. and Myers, R.M.	
JOURNAL		Expressed sequence tags from Gasterosteus aculeatus (2004)	
COMMENT		Unpublished (2004) Contact: Grimwood, Jane Stanford Human Genome Center Stanford University School of Medicine 975 S California Ave, Palo Alto, CA 94304, USA Tel: 650 320 5917 Fax: 650 320 5801 Email: jane@shgc.stanford.edu Plate: 362 High quality sequence stop: 888.	
FEATURES		Location/Qualifiers	
source		1. .906	
		/organism="Gasterosteus aculeatus"	
		/mol_type="mRNA"	
		/strain="Conner Creek sticklebacks, WA USA"	
		/db_xref="taxon:69293"	
		/clone="CNB362-G10"	
		/sex="mixed male and female"	
		/tissue_type="brain"	
		/dev_stage="adult"	
		/lab_host="DH10B (T1 phage resistant)"	
		/clone_lib="SHGC-CNB2"	
		/note="Vector: Express 1; Total and poly A+ RNA was isolated from the indicated stickleback tissue, and a cDNA library was constructed in the Express 1 plasmid vector by Open Biosystems. First strand cDNA synthesis was primed with an 54 bp linker primer containing an oligodT sequence preceded by a synthetic NotI site (first strand primer: 5'-GACTAGTTAGATCGAGCGCGCC(T)25-3'). Following second strand synthesis, cDNAs were made blunt at the end corresponding to the original 5 prime end of mRNA, and cloned directionally into the NotI and EcoRV sites of Express 1. Note that the EcoRV site is typically destroyed in the blunt end cloning, leaving a junction of the form 'xxxATC' (where is ATC is the second half of the EcoRV site, and xxx is derived from the cDNA sequence). A map of the Express 1 vector is available at: http://www.openbiosystems.com/cdna_library.constructionfaq.php#8 The primary library was transformed and amplified in DH10B (T1 phage resistant) bacteria."	
ORIGIN			
Query Match		62.7%; Score 20.8; DB 10; Length 906;	
Best Local Similarity		52.4%; Pred. No. 8.2e+02;	
Matches		22; Conservative 6; Mismatches 14; Indels 0; Gaps 0;	
QY		1 GCNCAYGNTGCCNTCNCNTAYAAAGAYTAGGTNAAR 42	
Db		160 GCCCGCGTGGTCTTCGCCGTTCAACGATTCAAAGAGAA 119	

```

Db      685 GCCCATGTTGGATTCTCGTACAGAAAAATGAAATAAAA 726

RESULT 95
LOCUS   CB209690
DEFINITION AGENCOURT 11302488 NICHG XGC Emb1 Xenopus laevis cDNA clone
IMAGE: 6865071 5', mRNA sequence.
ACCESSION CB209690
VERSION   CB209690.1 GI:28251253
KEYWORDS EST.
SOURCE    Xenopus laevis (African clawed frog)
ORGANISM  Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
Xenopodinae; Xenopus; Xenopus.
REFERENCE 1 (bases 1 to 967)
AUTHORS   NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
Email: cgapsb-remail.nih.gov
Tissue Procurement: Martha Rebbert, Steven L. Klein, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: NCI-CCGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM14486 row: n column: 14
High quality sequence stop: 311.
Location/Qualifiers
FEATURES             source
    1..967
    /organism="Xenopus laevis"
    /mol_type="mRNA"
    /db_xref="taxon:8355"
    /clone="IMAGE:6865071"
    /tissue_type="embryo (stage 10)"
    /lab_host="DH10B (phage-resistant)"
    /clone_lib="NICHG XGC Emb1"
    /notes="Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI;
Cloned unidirectionally. Primer: Oligo dT. Average insert
size 1.55 kb. Constructed by Life Technologies. Note: This
is a Xenopus Gene Collection (XGC) library."

ORIGIN
Query Match      62.7%; Score 20.8; DB 4; Length 967;
Best Local Similarity 52.4%; Pred. No. 8.3e+02;
Matches 22; Conservative 6; Mismatches 14; Indels 0; Gaps 0;

QY      1 GCNCAVCNTGCCNTCCNCTAYAGAYTAYGAGTNAAR 42
        ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      171 GCTCATGCTTGGAAACCTGGAACCAAGATGCGAAGTAAAG 212

RESULT 96
LOCUS   CV922581/c
DEFINITION PHRCH 2055 cysts, germinating Phytophthora infestans cDNA, mRNA
sequence.
ACCESSION CV922581
VERSION   CV922581.1 GI:58112124
KEYWORDS EST.
SOURCE    Phytophthora infestans (potato late blight agent)
ORGANISM  Phytophthora infestans
Eukaryota; stramenopiles; Oomycetes; Peronosporales; Phytophthora.
REFERENCE 1 (bases 1 to 974)
AUTHORS   Randall, T., Dwyer, R.A., Huitema, E., Beyer, K., Cvitanich, C.,
Kelkar, H., Fong, A.M., Gates, K., Roberts, S., Yatskan, E.,
Law, M., Testa, A., Torto-Alalibo, A., Zhang, M., Zheng, L., Mueller, E.,
Windass, J., Binder, A., Birch, P.R.J., Gisi, U., Govers, F., Gow, N.A.,
Mauch, F., van West, P., Waugh, M.E., Yu, J., Boller, T., Kamoun, S.,

Lam, S.T. and Judelson, H.S.
Large-scale gene discovery in the oomycete Phytophthora infestans
reveals likely components of phytopathogenicity shared with true
fungi
Mol. Plant-Microbe Interact. 18 (3), 229-243 (2005)
15782637
Contact: Judelson HS
Department of Plant Pathology
University of California
Webber Hall, Riverside, CA 92521, USA
Tel: 909 787 4199
Fax: 909 787 4294
Email: howard.judelson@ucr.edu.
Location/Qualifiers
FEATURES             source
    1..974
    /organism="Phytophthora infestans"
    /mol_type="mRNA"
    /strain="88069"
    /db_xref="taxon:4787"
    /sex="Al"
    /clone_lib="cysts, germinating"
    /note="Vector: pSPORT1"

ORIGIN
Query Match      62.7%; Score 20.8; DB 8; Length 974;
Best Local Similarity 52.4%; Pred. No. 8.3e+02;
Matches 22; Conservative 6; Mismatches 14; Indels 0; Gaps 0;

QY      1 GCNCAVCNTGCCNTCCNCTAYAGAYTAYGAGTNAAR 42
        ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      878 GCACAAACTGACCATCTCGTATCGAGATTTGTGTCAG 837

RESULT 97
LOCUS   CD242359/c
DEFINITION AGENCOURT 14122032 NIH MGC 179 Homo sapiens cDNA clone
IMAGE: 30383924 5', mRNA sequence.
ACCESSION CD242359
VERSION   CD242359.1 GI:31002823
KEYWORDS EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE 1 (bases 1 to 1047)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
Email: cgapsb-remail.nih.gov
Tissue Procurement: Dr. Michael Brownstein
cDNA Library Preparation: Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: NDAM448 row: i column: 21
High quality sequence stop: 453.
Location/Qualifiers
FEATURES             source
    1..1047
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone="IMAGE:30383924"
    /tissue_type="Pituitary"
    /lab_host="DH10B-ron A ( T1 and T5 phage resistances)"
    /clone_lib="NIH MGC 179"
    /notes="Organ: brain; Vector: pCMV-SPORT6.1; Site 1: EcoRV
(destroyed); Site 2: NotI; Library is oligo-dT primed and
directionally cloned (EcoRV site is destroyed upon

```



LOCUS DN670301 1226 bp mRNA linear EST 29-MAR-2005  
DEFINITION CFW55-G12.Y1d-s SHGC-CFW Gasterosteus aculeatus cDNA clone  
ACCESSION DN670301 CFW55-G12 5', mRNA sequence.  
VERSION DN670301.1 GI:61990358  
KEYWORDS EST.  
SOURCE Gasterosteus aculeatus (three spined stickleback)  
ORGANISM Gasterosteus aculeatus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes;  
Gasterosteidae; Gasterosteus.  
REFERENCE 1 (bases 1 to 1226)  
AUTHORS Kingley,D.M., Peichel,C., Balabhadra,S., Grimwood,J., Dickson,M.,  
Schmutz,J. and Myers,R.M.  
TITLE Expressed sequence tags from Gasterosteus aculeatus  
JOURNAL Unpublished (2003)  
COMMENT Contact: Grimwood, Jane  
Stanford Human Genome Center  
Stanford University School of Medicine  
975 S California Ave, Palo Alto, CA 94304, USA  
Tel: 650 320 5917  
Fax: 650 320 5801  
Email: jane@shgc.stanford.edu  
Plate: 55  
High quality sequence start: 18  
High quality sequence stop: 898.  
FEATURES  
source Location/Qualifiers  
1..1226  
/organism="Gasterosteus aculeatus"  
/mol\_type="mRNA"  
/strain="Conner Creek sticklebacks, WA USA"  
/db\_xref="taxon:69293"  
/clone="CFW55-G12"  
/sex="mixed male and female"  
/tissue\_type="gills"  
/dev\_stage="adult"  
/lab\_host="DH10B (T1 phage resistant)"  
/clone\_lib="SHGC-CFW"  
/note="Vector: Express 1; Total and poly A+ RNA was  
isolated from the indicated stickleback tissue, and a cDNA  
library was constructed in the Express 1 plasmid vector by  
Open Biosystems. First strand cDNA synthesis was primed  
with an 54 bp linker primer containing an oligodT sequence  
preceded by a synthetic NotI site (first strand primer:  
5'-GACTAGTTTGTAGTCGCGAGCGCGCCCTT)25-3'). Following  
second strand synthesis, cDNAs were made blunt at the end  
corresponding to the original 5 prime end of mRNA, and  
cloned directionally into the NotI and EcoRV sites of  
Express 1. Note that the EcoRV site is typically destroyed  
in the blunt end cloning, leaving a junction of the form  
'xxATC' (where ATC is the second half of the EcoRV  
site, and xxx is derived from the cDNA sequence). A map of  
the Express 1 vector is available at:  
<http://www.openbiosystems.com/cdna> library construction fa  
q.php#8 The primary library was transformed and amplified  
in DH10B (T1 phage resistant) bacteria. Clones available  
from Open Biosystems:  
<http://www.openbiosystems.com/stickleback>"

ORIGIN  
Query Match 62.7%; Score 20.8; DB 9; Length 1226;  
Best Local Similarity 52.4%; Pred.No. 8.7e+02;  
Matches 22; Conservative 6; Mismatches 14; Indels 0; Gaps 0;  
QY 1 GCNCGCMTGCCNTCCNTAYAGAYTAYGAGTNAAR 42  
|||:|||||:|||||:|||||:|||||:|||||:  
Db 727 GCCCATGTTTGATTTCTCGTACAGAAAAATGAATRAAA 768

Search completed: June 10, 2006, 19:36:16  
Job time : 5235.8 secs

**THIS PAGE IS BLANK**



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2006, 15:21:38 ; Search time 94.15 Seconds  
(without alignments)  
834.695 Million cell updates/sec

Title: US-10-600-816-33

Perfect score: 33.2

Sequence: 1 gcnaygntgcctncnc.....ayaargaytaygargtarnaar 42

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1403666 seqs, 935554401 residues

Total number of hits satisfying chosen parameters: 2807332

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 1000 summaries

Database :

Issued Patents NA:\*

1: /EMC Celerra\_SIDS3/ptodata/2/ina/1 COMB.seq:\*

2: /EMC Celerra\_SIDS3/ptodata/2/ina/5 COMB.seq:\*

3: /EMC Celerra\_SIDS3/ptodata/2/ina/6A COMB.seq:\*

4: /EMC Celerra\_SIDS3/ptodata/2/ina/6B COMB.seq:\*

5: /EMC Celerra\_SIDS3/ptodata/2/ina/7 COMB.seq:\*

6: /EMC Celerra\_SIDS3/ptodata/2/ina/H COMB.seq:\*

7: /EMC Celerra\_SIDS3/ptodata/2/ina/PCTUS COMB.seq:\*

8: /EMC Celerra\_SIDS3/ptodata/2/ina/PP COMB.seq:\*

9: /EMC Celerra\_SIDS3/ptodata/2/ina/RE COMB.seq:\*

10: /EMC Celerra\_SIDS3/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	30	90.4	601	3	US-09-949-016-29173
2	30	90.4	601	3	US-09-949-016-126957
3	30	90.4	1212	3	US-09-188-930-249
4	30	90.4	1212	3	US-09-312-283C-249
5	30	90.4	1228	3	US-09-495-050A-223
6	30	90.4	2289	3	US-09-949-016-3613
7	30	90.4	2456	3	US-09-949-016-572
8	30	90.4	26086	3	US-09-949-016-15355
9	30	90.4	26238	3	US-09-949-016-12314
10	20.4	61.4	1053	4	US-09-605-703B-1879
11	20	60.2	285986	3	US-09-949-016-12287
12	20	60.2	1230025	3	US-09-198-452A-1
13	20	60.2	1230230	3	US-09-438-185A-1
14	19.6	59.0	1005	3	US-09-248-796A-3066
15	19.4	58.4	2416	3	US-09-799-451-392
16	19.4	58.4	3738	3	US-09-543-681A-1420
17	19.2	57.8	1001	3	US-09-641-638-394
18	19.2	57.8	1001	3	US-10-170-097-394
19	19.2	57.8	1828	3	US-10-197-220-6
20	19.2	57.8	2496	3	US-09-976-594-637
21	19.2	57.8	10627	2	US-08-060-925A-12
22	19.2	57.8	12222	3	US-09-328-925-42
23	19.2	57.8	18141	3	US-09-949-016-14063

19	57.2	184	3	US-09-513-999C-21553	Sequence 21553, A
25	56.0	400	3	US-08-956-171B-2945	Sequence 2945, Ap
26	56.0	400	3	US-08-781-986A-2945	Sequence 2945, Ap
27	56.0	1305	3	US-09-540-236-933	Sequence 933, App
28	56.0	8190	3	US-09-949-016-17412	Sequence 17412, A
29	56.0	94750	3	US-09-596-002-38	Sequence 38, Appl
30	56.0	111677	3	US-09-949-016-16946	Sequence 16946, A
31	56.0	129908	3	US-09-585-858-1	Sequence 1, Appli
32	56.0	129908	3	US-10-270-878-1	Sequence 1, Appli
33	55.4	333	3	US-09-270-767-30495	Sequence 30495, A
34	55.4	1113	3	US-09-270-767-14344	Sequence 14344, A
35	55.4	6450	3	US-09-949-016-16596	Sequence 16596, A
36	55.4	13908	3	US-09-949-016-16726	Sequence 16726, A
37	55.4	19698	3	US-09-949-016-14079	Sequence 14079, A
38	55.4	23155	3	US-09-949-016-12366	Sequence 12366, A
39	55.4	35058	3	US-09-949-016-12607	Sequence 12607, A
40	55.4	35059	3	US-09-949-016-13831	Sequence 13831, A
41	55.4	45571	3	US-09-949-016-16262	Sequence 16262, A
42	55.4	81001	3	US-09-750-580-1	Sequence 1, Appli
43	55.4	209631	3	US-09-949-002-574	Sequence 574, App
44	55.4	209632	3	US-09-949-002-802	Sequence 802, App
45	55.4	260247	3	US-09-949-016-13358	Sequence 13358, A
46	54.8	171	3	US-08-991-789A-60	Sequence 60, Appl
47	54.8	171	3	US-09-062-451-60	Sequence 60, Appl
48	54.8	171	3	US-09-598-326-60	Sequence 60, Appl
49	54.8	171	3	US-09-289-198-60	Sequence 60, Appl
50	54.8	171	3	US-09-429-755-60	Sequence 60, Appl
51	54.8	171	3	US-09-699-295-60	Sequence 60, Appl
52	54.8	171	3	US-09-534-825A-60	Sequence 60, Appl
53	54.8	601	3	US-09-949-016-107890	Sequence 107890, A
54	54.8	669	3	US-09-302-540-1562	Sequence 1562, Ap
55	54.8	864	3	US-09-328-352-1269	Sequence 1269, Ap
56	54.8	1101	3	US-09-328-352-1336	Sequence 1336, Ap
57	54.8	1335	3	US-09-134-000C-2009	Sequence 2009, Ap
58	54.8	45418	3	US-09-949-016-15870	Sequence 15870, A
59	54.8	183770	3	US-09-949-016-15494	Sequence 15494, A
60	54.8	374159	3	US-09-949-016-15868	Sequence 15868, A
61	54.8	390890	3	US-09-949-016-14720	Sequence 14720, A
62	54.2	541	3	US-09-533-559-1061	Sequence 1061, Ap
63	54.2	601	3	US-09-949-016-151014	Sequence 151014, A
64	54.2	601	3	US-09-949-016-193564	Sequence 193564, A
65	54.2	878	2	US-08-469-667-8	Sequence 8, Appli
66	54.2	878	3	US-09-224-110-8	Sequence 8, Appli
67	54.2	878	3	US-09-988-292A-8	Sequence 8, Appli
68	54.2	878	7	PCT-US95-07289-8	Sequence 8, Appli
69	54.2	1512	3	US-09-016-434-850	Sequence 850, App
70	54.2	1659	4	US-10-044-205A-1	Sequence 3, Appli
71	54.2	1662	3	US-09-738-894A-1	Sequence 1, Appli
72	54.2	1662	3	US-09-802-117-1	Sequence 1, Appli
73	54.2	1662	3	US-09-964-469-1	Sequence 1, Appli
74	54.2	1662	3	US-10-217-745-1	Sequence 1, Appli
75	54.2	1662	5	US-10-425-962-1	Sequence 1, Appli
76	54.2	1953	3	US-09-583-110-560	Sequence 560, App
77	54.2	1953	3	US-09-107-433-912	Sequence 912, App
78	54.2	2148	3	US-09-248-796A-837	Sequence 837, App
79	54.2	2198	4	US-10-044-205A-1	Sequence 1, Appli
80	54.2	2249	3	US-09-802-117-5	Sequence 5, Appli
81	54.2	2249	3	US-10-217-745-5	Sequence 5, Appli
82	54.2	2371	2	US-08-757-541-4	Sequence 4, Appli
83	54.2	2371	3	US-09-033-275-4	Sequence 4, Appli
84	54.2	2371	3	US-09-342-581-4	Sequence 4, Appli
85	54.2	2745	3	US-09-623-624-5	Sequence 5, Appli
86	54.2	2745	3	US-10-270-595-5	Sequence 5, Appli
87	54.2	2763	3	US-09-685-403-1	Sequence 1, Appli
88	54.2	3007	3	US-09-193-562D-27	Sequence 27, Appl
89	54.2	3007	3	US-10-055-412B-27	Sequence 27, Appl
90	54.2	7172	3	US-08-961-527-120	Sequence 120, App
91	54.2	7411	3	US-10-001-189-67	Sequence 67, Appl
92	54.2	15566	3	US-09-949-016-17104	Sequence 17104, A
93	54.2	36851	3	US-09-738-894A-3	Sequence 3, Appli
94	54.2	36851	3	US-09-964-469-3	Sequence 3, Appli
95	54.2	36851	5	US-10-425-962-3	Sequence 3, Appli
96	54.2	100550	3	US-09-949-016-11835	Sequence 11835, A















c 973 16.4 49.4 818128 3 US-09-949-016-14547  
c 974 16.4 49.4 818128 3 US-09-949-016-14548  
c 975 16.4 49.4 818128 3 US-09-949-016-14549  
c 976 16.4 49.4 818128 3 US-09-949-016-14550  
c 977 16.4 49.4 818128 3 US-09-949-016-14551  
c 978 16.4 49.4 818128 3 US-09-949-016-14552  
c 979 16.4 49.4 818128 3 US-09-949-016-14553  
c 980 16.4 49.4 818128 3 US-09-949-016-14554  
c 981 16.4 49.4 818128 3 US-09-949-016-14555  
c 982 16.4 49.4 818128 3 US-09-949-016-14556  
c 983 16.4 49.4 818128 3 US-09-949-016-14557  
c 984 16.4 49.4 818128 3 US-09-949-016-14558  
c 985 16.4 49.4 818128 3 US-09-949-016-14559  
c 986 16.4 49.4 818128 3 US-09-949-016-14560  
c 987 16.4 49.4 818128 3 US-09-949-016-14561  
c 988 16.4 49.4 818128 3 US-09-949-016-14562  
c 989 16.4 49.4 818128 3 US-09-949-016-14563  
c 990 16.4 49.4 818128 3 US-09-949-016-14564  
c 991 16.4 49.4 818128 3 US-09-949-016-14565  
c 992 16.4 49.4 818128 3 US-09-949-016-14566  
c 993 16.4 49.4 1082144 4 US-09-531-120-211  
c 994 16.2 48.8 89 3 US-09-254-126D-22  
c 995 16.2 48.8 234 3 US-09-540-236-1801  
c 996 16.2 48.8 285 3 US-09-543-681A-2616  
c 997 16.2 48.8 288 3 US-08-943-731-55  
c 998 16.2 48.8 290 5 US-09-974-300-8070  
c 999 16.2 48.8 346 3 US-09-621-976-13443  
1000 16.2 48.8 351 3 US-09-621-976-13420, A

## ALIGNMENTS

## RESULT 1

US-09-949-016-29173  
; Sequence 29173, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 29173  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-29173

Query Match 90.4%; Score 30; DB 3; Length 601;  
Best Local Similarity 64.3%; Pred. No. 0.00071;  
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 GNCAYGNTGGCNCNTAYAAAGAYTAYGAGTNAAR 42  
Db 14 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 55

## RESULT 2

US-09-949-016-126957  
; Sequence 126957, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.

Sequence 14547, A  
Sequence 14548, A  
Sequence 14549, A  
Sequence 14550, A  
Sequence 14551, A  
Sequence 14552, A  
Sequence 14553, A  
Sequence 14554, A  
Sequence 14555, A  
Sequence 14556, A  
Sequence 14557, A  
Sequence 14558, A  
Sequence 14559, A  
Sequence 14560, A  
Sequence 14561, A  
Sequence 14562, A  
Sequence 14563, A  
Sequence 14564, A  
Sequence 14565, A  
Sequence 14566, A  
Sequence 14567, A  
Sequence 211, Appl  
Sequence 22, Appl  
Sequence 1801, Ap  
Sequence 2616, Ap  
Sequence 55, Appl  
Sequence 8070, Ap  
Sequence 13443, A  
Sequence 13420, A

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 126957  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-126957

Query Match 90.4%; Score 30; DB 3; Length 601;  
Best Local Similarity 64.3%; Pred. No. 0.00071;  
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 GNCAYGNTGGCNCNTAYAAAGAYTAYGAGTNAAR 42  
Db 14 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 55

## RESULT 3

US-09-188-930-249  
; Sequence 249, Application US/09188930A  
; Patent No. 6150502  
; GENERAL INFORMATION:  
; APPLICANT: Watson, James D.  
; APPLICANT: Strachan, Lorna  
; APPLICANT: Sleeman, Matthew  
; APPLICANT: Onrust, Rene  
; APPLICANT: Murison, James Greg  
; TITLE OF INVENTION: Compositions Isolated From Skin Cells  
; TITLE OF INVENTION: and Methods For Their Use  
; FILE REFERENCE: 11000.1011c1  
; CURRENT APPLICATION NUMBER: US/09/188,930A  
; CURRENT FILING DATE: 1998-11-09  
; NUMBER OF SEQ ID NOS: 348  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 249  
; LENGTH: 1212  
; TYPE: DNA  
; ORGANISM: Human  
US-09-188-930-249

Query Match 90.4%; Score 30; DB 3; Length 1212;  
Best Local Similarity 64.3%; Pred. No. 0.00089;  
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 GNCAYGNTGGCNCNTAYAAAGAYTAYGAGTNAAR 42  
Db 990 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1031

## RESULT 4

US-09-312-283C-249  
; Sequence 249, Application US/09312283C  
; Patent No. 6573095  
; GENERAL INFORMATION:  
; APPLICANT: Watson, James D.  
; APPLICANT: Strachan, Lorna  
; APPLICANT: Sleeman, Matthew  
; APPLICANT: Onrust, Rene  
; APPLICANT: Murison, James G.  
; APPLICANT: Kumble, Krishanand D.  
; TITLE OF INVENTION: Compositions Isolated from Skin Cells  
; TITLE OF INVENTION: and Methods for Their Use



; FILE REFERENCE: 11000.1011c2  
; CURRENT APPLICATION NUMBER: US/09/312,283C  
; CURRENT FILING DATE: 1999-05-14  
; NUMBER OF SEQ ID NOS: 425  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 249  
; LENGTH: 1212  
; TYPE: DNA  
; ORGANISM: Mouse  
US-09-312-283C-249

Query Match 90.4%; Score 30; DB 3; Length 1212;  
Best Local Similarity 64.3%; Pred. No. 0.00089;  
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

Qy 1 GCNCAYCNCNTGCCNTCCNTAYAAAGAYTAYGARGCTNAAR 42  
Db 990 GCCACGCTTGGCGGAGCCCTTACAAAGACTATGAAGTAAAG 1031

## RESULT 5

US-09-495-050A-223  
; Sequence 223, Application US/09495050A  
; Patent No. 6492505

; GENERAL INFORMATION:  
; APPLICANT: Roopa, Reddy

; APPLICANT: Guegler, Karl, J.  
; APPLICANT: Au-Young, Janice

; TITLE OF INVENTION: COMPOSITION FOR DETECTION OF GENES ENCODING MEMBRANE-ASSOCIATED  
; FILE REFERENCE: PA-0013 US

; CURRENT APPLICATION NUMBER: US/09/495,050A  
; CURRENT FILING DATE: 2000-01-31

; PRIOR APPLICATION NUMBER: 60/118,318  
; PRIOR FILING DATE: February 1, 1999

; NUMBER OF SEQ ID NOS: 305  
; SOFTWARE: PERL Program

; SEQ ID NO 223  
; LENGTH: 1228  
; TYPE: DNA

; ORGANISM: Homo sapiens  
; NAME/KEY: misc feature

; OTHER INFORMATION: Incyte ID No. 6492505 2681738CT1  
US-09-495-050A-223

Query Match 90.4%; Score 30; DB 3; Length 1228;  
Best Local Similarity 64.3%; Pred. No. 0.00089;  
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

Qy 1 GCNCAYCNCNTGCCNTCCNTAYAAAGAYTAYGARGCTNAAR 42  
Db 541 GCCACGCTTGGCGGAGCCCTTACAAAGACTATGAAGTAAAG 582

## RESULT 6

US-09-495-016-3613  
; Sequence 3613, Application US/09949016  
; Patent No. 6812339

; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755

; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768

; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498

; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 3613  
; LENGTH: 2289  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-3613

Query Match 90.4%; Score 30; DB 3; Length 2289;  
Best Local Similarity 64.3%; Pred. No. 0.0011;  
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

Qy 1 GCNCAYCNCNTGCCNTCCNTAYAAAGAYTAYGARGCTNAAR 42  
Db 1117 GCCACGCTTGGCGGAGCCCTTACAAAGACTATGAAGTAAAG 1158

## RESULT 7

US-09-949-016-572  
; Sequence 572, Application US/09949016  
; Patent No. 6812339

; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755

; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768

; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498

; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 572  
; LENGTH: 2456  
; TYPE: DNA

; ORGANISM: Human  
US-09-949-016-572

Query Match 90.4%; Score 30; DB 3; Length 2456;  
Best Local Similarity 64.3%; Pred. No. 0.0011;  
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

Qy 1 GCNCAYCNCNTGCCNTCCNTAYAAAGAYTAYGARGCTNAAR 42  
Db 1271 GCCACGCTTGGCGGAGCCCTTACAAAGACTATGAAGTAAAG 1312

## RESULT 8

US-09-949-016-15355  
; Sequence 15355, Application US/09949016  
; Patent No. 6812339

; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755

; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768

; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498

; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 15355  
; LENGTH: 26086  
; TYPE: DNA

; ORGANISM: Human  
US-09-949-016-15355



```
LOCATION: (30001)..(45000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (405001)..(420000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (450001)..(600000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (600001)..(750000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (750001)..(900000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (900001)..(1050000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (1050001)..(1200000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (1200001)..(1350000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (1350001)..(1500000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (1500001)..(1650000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (1650001)..(1800000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (1800001)..(1950000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (1950001)..(2100000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (2100001)..(2250000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (2250001)..(2400000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (2400001)..(2550000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (2550001)..(2700000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (2700001)..(2850000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (2850001)..(3000000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (3000001)..(3150000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (3150001)..(3300000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (3300001)..(3450000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (3450001)..(3600000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (3600001)..(3750000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (3750001)..(3900000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (3900001)..(4050000)
```



```
; Query Match 58.4%; Score 19.4; DB 3; Length 2416;
; Best Local Similarity 55.6%; Pred. No. 85;
; Matches 20; Conservative 5; Mismatches 11; Indels 0; Gaps 0;

Qy 1 GCNAYCNCNTGCCNTCCNTAYARGATYAGAR 36
Db 1965 GCACACGATGGCTTGGCCAGACACAAACCAAGAG 1930

RESULT 16
US-09-543-681A-1420
; Sequence 1420, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 1420
; LENGTH: 3738
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-1420

Query Match 58.4%; Score 19.4; DB 3; Length 3738;
; Best Local Similarity 52.6%; Pred. No. 97;
; Matches 20; Conservative 6; Mismatches 12; Indels 0; Gaps 0;

Qy 5 AYCNTGCCNTCCNTAYARGATYAGCTNAR 42
Db 3278 AAGCATGGATTGCACACTTCAAGATTACAAAGTGAA 3315

RESULT 17
US-09-641-638-394/c
; Sequence 394, Application US/09641638
; Patent No. 6432648
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Chumakov, Ilya
; APPLICANT: Cohen, Annick
; TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
; FILE REFERENCE: GENSET 051CP1
; CURRENT APPLICATION NUMBER: US/09/641,638
; CURRENT FILING DATE: 2000-08-16
; PRIOR APPLICATION NUMBER: US 09/502,330
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: US 60/133,200
; PRIOR FILING DATE: 1999-05-07
; PRIOR APPLICATION NUMBER: US 09/275,267
; PRIOR FILING DATE: 1999-03-23
; PRIOR APPLICATION NUMBER: US 60/119,917
; NUMBER OF SEQ ID NOS: 1304
; SOFTWARE: Patent.pm
; SEQ ID NO 394
; LENGTH: 1001
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: allele
; LOCATION: 479
; OTHER INFORMATION: 12-889-518 : polymorphic base G or A
; FEATURE:
; NAME/KEY: misc_binding
; LOCATION: 459..478
; OTHER INFORMATION: 12-889-518.misl, potential
; FEATURE:
; NAME/KEY: misc_binding
; LOCATION: 480..499
; OTHER INFORMATION: 12-889-518.mis2, potential complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 976..995
; OTHER INFORMATION: upstream amplification primer, complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 426..445
; OTHER INFORMATION: downstream amplification primer

Qy 8 CNTGGCCNTCCNTAYARGATYAGA 35
Db 169 CATGCCATCTTCTTATAGGACAATGA 142

RESULT 18
US-10-170-097-394/c
; Sequence 394, Application US/10170097
; Patent No. 6794143
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Chumakov, Ilya
; APPLICANT: Cohen, Annick
; TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
; FILE REFERENCE: GEN-T14XC2D1
; CURRENT APPLICATION NUMBER: US/10/170,097
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: US 09/641,638
; PRIOR FILING DATE: 2000-08-16
; PRIOR APPLICATION NUMBER: US 09/502,330
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: US 60/133,200
; PRIOR FILING DATE: 1999-05-07
; PRIOR APPLICATION NUMBER: US 09/275,267
; PRIOR FILING DATE: 1999-03-23
; PRIOR APPLICATION NUMBER: US 60/119,917
; NUMBER OF SEQ ID NOS: 1304
; SOFTWARE: Patent.pm
; SEQ ID NO 394
; LENGTH: 1001
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: allele
; LOCATION: 479
; OTHER INFORMATION: 12-889-518 : polymorphic base G or A
; FEATURE:
; NAME/KEY: misc_binding
; LOCATION: 459..478
; OTHER INFORMATION: 12-889-518.misl, potential
; FEATURE:
; NAME/KEY: misc_binding
; LOCATION: 480..499
; OTHER INFORMATION: 12-889-518.mis2, potential complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 976..995
; OTHER INFORMATION: upstream amplification primer, complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 426..445
; OTHER INFORMATION: downstream amplification primer
```

```
; NAME/KEY: misc_binding
; LOCATION: 480..499
; OTHER INFORMATION: 12-889-518.mis2, potential complement
; NAME/KEY: primer_bind
; LOCATION: 976..995
; OTHER INFORMATION: upstream amplification primer, complement
; NAME/KEY: primer_bind
; LOCATION: 426..445
; OTHER INFORMATION: downstream amplification primer
; NAME/KEY: misc_binding
; LOCATION: 467..491
; OTHER INFORMATION: 12-889-518 potential probe
US-09-641-638-394
```

```
Query Match 57.8%; Score 19.2; DB 3; Length 1001;
; Best Local Similarity 64.3%; Pred. No. 80;
; Matches 18; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Qy 8 CNTGGCCNTCCNTAYARGATYAGA 35
Db 169 CATGCCATCTTCTTATAGGACAATGA 142
```



```
; SEQ ID NO 42
; LENGTH: 12222
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-328-925-42

Query Match      57.8%; Score 19.2; DB 3; Length 12222;
Best Local Similarity 50.0%; Pred. No. 1.7e+02;
Matches 21; Conservative 6; Mismatches 15; Indels 0; Gaps 0;

QY 1 GCNCAYCNCNTGCCNTCNCNTAYAARGAYTAYGARGTNAAR 42
Db 2432 GGCATGCATCTTTCTCCCTTCTATAGATGAGGAGGTTAAG 2473

RESULT 23
US-09-949-016-14063/c
; Sequence 14063, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-03
; PRIOR FILING DATE: 2000-10-03
; PRIOR FILING DATE: 2000-09-08
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14063
; LENGTH: 18141
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(18141)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14063

Query Match      57.8%; Score 19.2; DB 3; Length 18141;
Best Local Similarity 56.8%; Pred. No. 2e+02;
Matches 21; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

QY 5 AYCNTGGCCNTCNCNTAYAARGAYTAYGARGTNA 41
Db 12619 ACGCTGGACTTCTCAAAATTATGAGCATGAGTCAA 12583

RESULT 24
US-09-513-999C-21553/c
; Sequence 21553, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59 US2,REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 21553
; LENGTH: 184
; TYPE: DNA

; ORGANISM: Homo sapiens
US-09-513-999C-21553

Query Match      57.2%; Score 19; DB 3; Length 184;
Best Local Similarity 63.3%; Pred. No. 58;
Matches 19; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 1 GCNCAYCNCNTGCCNTCNCNTAYAARGAY 30
Db 44 GCGCAAGCCTGGCCTTCCCCAACAACAGAT 15

RESULT 25
US-08-956-171E-2945
; Sequence 2945, Application US/08956171E
; Patent No. 6593114
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; APPLICANT: Gil H. Choi
; APPLICANT: Patrick S. Dillon
; APPLICANT: Craig A. Rosen
; APPLICANT: Steven C. Barash
; APPLICANT: Michael R. Fannon
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5256
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/956,171E
; FILING DATE: 20-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/009,861
; FILING DATE: January 5, 1996
; APPLICATION NUMBER: 08/781,986
; FILING DATE: January 3, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mark J. Hyman
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB248P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (240) 314-1224
; TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 2945:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 400 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 2945:
US-08-956-171E-2945

Query Match      56.0%; Score 18.6; DB 3; Length 400;
Best Local Similarity 51.2%; Pred. No. 1.1e+02;
Matches 21; Conservative 5; Mismatches 15; Indels 0; Gaps 0;

QY 1 GCNCAYCNCNTGCCNTCNCNTAYAARGAYTAYGARGTNA 41
Db 354 GCACTGGCTGGCCAACTCGTTTTCAGGATTATCATCCAAA 394

RESULT 26
US-08-781-986A-2945
```





```
US-09-949-016-16946/c
; Sequence 16946, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CU001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16946
; LENGTH: 111677
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16946

Query Match 56.0%; Score 18.6; DB 3; Length 111677;
Best Local Similarity 56.2%; Pred. No. 6.3e+02;
Matches 18; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Qy 2 CNCAYGNTGCCNTCCNTAYAAARGAYTAY 33
Db 47626 CCCATGCCGGCCCTGCCCAAGACTTAC 47595

RESULT 31
US-09-585-858-1
; Sequence 1, Application US/09585858
; Patent No. 6492161
; GENERAL INFORMATION:
; APPLICANT: Sigrður Hjørleifsdóttir
; APPLICANT: Gudmundur O. Hreggvidsson
; APPLICANT: Olafur H. Fridjonsson
; APPLICANT: Arthor Aevasson
; APPLICANT: Jakob K. Kristjansson
; TITLE OF INVENTION: Bacteriophage RM378 of a Thermophilic
; TITLE OF INVENTION: Host Organism
; FILE REFERENCE: 2739.1001-001
; CURRENT APPLICATION NUMBER: US/09/585,858
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: 60/137,120
; PRIOR FILING DATE: 1999-06-02
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 129908
; TYPE: DNA
; ORGANISM: Bacteriophage RM378
US-09-585-858-1

Query Match 56.0%; Score 18.6; DB 3; Length 129908;
Best Local Similarity 58.1%; Pred. No. 6.6e+02;
Matches 18; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

Qy 10 TGGCCNTCCNTAYAAARGAYTAYGARGTNA 40
Db 28468 TGGCTCTCACCTTTCAAAGATGATGAAGTGA 28498

RESULT 32
US-10-270-878-1
; Sequence 1, Application US/10270878
; Patent No. 6818425
; GENERAL INFORMATION:
; APPLICANT: Sigrður Hjørleifsdóttir
```

```
; APPLICANT: Gudmundur O. Hreggvidsson
; APPLICANT: Olafur H. Fridjonsson
; APPLICANT: Arthor Aevasson
; APPLICANT: Jakob K. Kristjansson
; TITLE OF INVENTION: Bacteriophage RM378 of a Thermophilic
; TITLE OF INVENTION: Host Organism
; FILE REFERENCE: 2739.1001-001
; CURRENT APPLICATION NUMBER: US/10/270,878
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: US/09/585,858
; PRIOR FILING DATE: 2000-12-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 129908
; TYPE: DNA
; ORGANISM: Bacteriophage RM378
US-10-270-878-1

Query Match 56.0%; Score 18.6; DB 3; Length 129908;
Best Local Similarity 58.1%; Pred. No. 6.6e+02;
Matches 18; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

Qy 10 TGGCCNTCCNTAYAAARGAYTAYGARGTNA 40
Db 28468 TGGCTCTCACCTTTCAAAGATGATGAAGTGA 28498

RESULT 33
US-09-270-767-30495/c
; Sequence 30495, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 30495
; LENGTH: 333
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-30495

Query Match 55.4%; Score 18.4; DB 3; Length 333;
Best Local Similarity 54.3%; Pred. No. 1.3e+02;
Matches 19; Conservative 5; Mismatches 11; Indels 0; Gaps 0;

Qy 2 CNCAYGNTGCCNTCCNTAYAAARGAYTAYGAR 36
Db 226 CGCATTGTTTGTGTCGCCGTATAAGACTAGGAA 192

RESULT 34
US-09-270-767-14344/c
; Sequence 14344, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14344
; LENGTH: 1113
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-14344
```

```
Query Match          55.4%; Score 18.4; DB 3; Length 1113;
Best Local Similarity 54.3%; Pred. No. 1.9e+02;
Matches 19; Conservative 5; Mismatches 11; Indels 0; Gaps 0;

QY      2  CNCAYCGTGGCCNTCNCNTAYAAARGAYTAYGAR 36
Db      1006  CGCATTCGTTGTGTGCGCGTATAAAGACTAGGAA 972

RESULT 35
US-09-949-016-16596
; Sequence 16596, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16596
; LENGTH: 6450
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16596

Query Match          55.4%; Score 18.4; DB 3; Length 6450;
Best Local Similarity 52.4%; Pred. No. 3.3e+02;
Matches 22; Conservative 4; Mismatches 16; Indels 0; Gaps 0;

QY      1  GCNCAYCGTGGCCNTCNCNTAYAAARGAYTAYGARGTNAAR 42
Db      1728  GCAAAAGCTTGGCCAGGCGCTTCAAGGATGGGAGGTACAG 1769

RESULT 36
US-09-949-016-16726
; Sequence 16726, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16726
; LENGTH: 13908
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16726

Query Match          55.4%; Score 18.4; DB 3; Length 13908;
Best Local Similarity 54.3%; Pred. No. 4.2e+02;
Matches 19; Conservative 5; Mismatches 11; Indels 0; Gaps 0;

QY      8  CNTGGCCNTCNCNTAYAAARGAYTAYGARGTNAAR 42
Db      1006  CGCATTCGTTGTGTGCGCGTATAAAGACTAGGAA 972

Query Match          55.4%; Score 18.4; DB 3; Length 1113;
Best Local Similarity 54.3%; Pred. No. 1.9e+02;
Matches 19; Conservative 5; Mismatches 11; Indels 0; Gaps 0;

QY      2  CNCAYCGTGGCCNTCNCNTAYAAARGAYTAYGAR 36
Db      1006  CGCATTCGTTGTGTGCGCGTATAAAGACTAGGAA 972

RESULT 37
US-09-949-016-14079/c
; Sequence 14079, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14079
; LENGTH: 19698
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14079

Query Match          55.4%; Score 18.4; DB 3; Length 19698;
Best Local Similarity 51.4%; Pred. No. 4.7e+02;
Matches 19; Conservative 6; Mismatches 12; Indels 0; Gaps 0;

QY      6  YGCTGGCCNTCNCNTAYAAARGAYTAYGARGTNAAR 42
Db      18964  TGGTGGACCTCCCTTTTAAAGATCTCTATGTATAA 18928

RESULT 38
US-09-949-016-12366
; Sequence 12366, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12366
; LENGTH: 23155
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(23155)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12366

Query Match          55.4%; Score 18.4; DB 3; Length 23155;
Best Local Similarity 52.4%; Pred. No. 4.9e+02;
Matches 22; Conservative 4; Mismatches 16; Indels 0; Gaps 0;

QY      1  GCNCAYCGTGGCCNTCNCNTAYAAARGAYTAYGARGTNAAR 42
Db      1006  CGCATTCGTTGTGTGCGCGTATAAAGACTAGGAA 972
```

Db 17278 GCAAAGCTTGGCCAGGCGCTTCAAGGATTGGAGGTACAG 17319

## RESULT 39

US-09-949-016-12607  
; Sequence 12607, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12607  
; LENGTH: 35058  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-12607

Query Match 55.4%; Score 18.4; DB 3; Length 35058;  
Best Local Similarity 48.7%; Pred. No. 5.6e+02;  
Matches 19; Conservative 7; Mismatches 13; Indels 0; Gaps 0;

QY 4 CAYGNTGGCCNTCNCNTAYARGAYTAYGARGTNAAR 42

Db 23763 CATACATCCTCCTCAGCTTACAAAGATGACGGGATTAA 23801

## RESULT 40

US-09-949-016-13831  
; Sequence 13831, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13831  
; LENGTH: 35059  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-13831

Query Match 55.4%; Score 18.4; DB 3; Length 35059;  
Best Local Similarity 48.7%; Pred. No. 5.6e+02;  
Matches 19; Conservative 7; Mismatches 13; Indels 0; Gaps 0;

QY 4 CAYGNTGGCCNTCNCNTAYARGAYTAYGARGTNAAR 42

Db 23763 CATACATCCTCCTCAGCTTACAAAGATGACGGGATTAA 23801

## RESULT 41

US-09-949-016-16262

; Sequence 16262, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 16262  
; LENGTH: 45571  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-16262

Query Match 55.4%; Score 18.4; DB 3; Length 45571;  
Best Local Similarity 59.4%; Pred. No. 6e+02;  
Matches 19; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 4 CAYGNTGGCCNTCNCNTAYARGAYTAYGA 35

Db 18161 CATGCTGGCCAATTCATATAAATTTAGGA 18192

## RESULT 42

US-09-750-580-1  
; Sequence 1, Application US/09750580  
; Patent No. 6455280  
; GENERAL INFORMATION:  
; APPLICANT: Yen, Frances  
; APPLICANT: Denison, Blake  
; APPLICANT: Bour, Barbara  
; APPLICANT: Bihain, Bernard  
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste  
; APPLICANT: Duclert, Aymeric  
; APPLICANT: Bougueleret, Lydie  
; APPLICANT: Ebbets-Reed, Dana  
; APPLICANT: Salter-Cid, Luisa  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITING NEOPLASTIC CELL GROWTH  
; FILE REFERENCE: 89 US2.CIP  
; CURRENT APPLICATION NUMBER: US/09/750,580  
; CURRENT FILING DATE: 2000-12-28  
; PRIOR APPLICATION NUMBER: US 09/599,362  
; PRIOR FILING DATE: 2000-06-21  
; PRIOR APPLICATION NUMBER: PCT/IB00/0101  
; PRIOR FILING DATE: 2000-06-21  
; PRIOR APPLICATION NUMBER: PCT/IB99/02058  
; PRIOR FILING DATE: 1999-12-20  
; PRIOR APPLICATION NUMBER: US 49/469/099  
; PRIOR FILING DATE: 1999-12-21  
; PRIOR APPLICATION NUMBER: US 60/113,686  
; PRIOR FILING DATE: 1998-12-22  
; PRIOR APPLICATION NUMBER: US 60/141,032  
; PRIOR FILING DATE: 1999-06-25  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: Patent.pm  
; SEQ ID NO 1  
; LENGTH: 81001  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 10946..12946  
; OTHER INFORMATION: 5'regulatory region  
; NAME/KEY: exon





```
; Patent No. 6344550
; GENERAL INFORMATION:
; APPLICANT: Frudakis, Tony N.
; APPLICANT: Smith, John M.
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF BREAST CANCER
; NUMBER OF SEQUENCES: 297
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 04-APR-1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.419C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 60:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 171 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-062-451-60

Query Match 54.8%; Score 18.2; DB 3; Length 171;
Best Local Similarity 52.6%; Pred. No. 1.3e+02;
Matches 20; Conservative 5; Mismatches 13; Indels 0; Gaps 0;

QY 5 AYCNTGGCCNTCCNTAARGAYTAYGARTNAAR 42
Db 134 ATGTGGCATTGACCATTAATACTCTAAGAGGTTAAA 97

RESULT 48
US-09-598-326-60/c
; Sequence 60, Application US/09598326
; Patent No. 6423496
; GENERAL INFORMATION:
; APPLICANT: Frudakis, Tony N.
; APPLICANT: Smith, John M.
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF BREAST CANCER
; NUMBER OF SEQUENCES: 247
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed Intellectual Property Law Group PLLC
; STREET: 701 Fifth Avenue, Suite 6300
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/598,326
```

```
; FILING DATE: 20-Jun-2000
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Potter, Jane E.R.
; REGISTRATION NUMBER: 33,332
; REFERENCE/DOCKET NUMBER: 210121.419D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 60:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 171 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 60:
US-09-598-326-60

Query Match 54.8%; Score 18.2; DB 3; Length 171;
Best Local Similarity 52.6%; Pred. No. 1.3e+02;
Matches 20; Conservative 5; Mismatches 13; Indels 0; Gaps 0;

QY 5 AYCNTGGCCNTCCNTAARGAYTAYGARTNAAR 42
Db 134 ATGTGGCATTGACCATTAATACTCTAAGAGGTTAAA 97

RESULT 49
US-09-289-198-60/c
; Sequence 60, Application US/09289198
; Patent No. 6586570
; GENERAL INFORMATION:
; APPLICANT: Frudakis, Tony N.
; APPLICANT: Smith, John M.
; APPLICANT: Reed, Steven G.
; APPLICANT: Misher, Lynda
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.419C5
; CURRENT APPLICATION NUMBER: US/09/289,198
; CURRENT FILING DATE: 1999-04-09
; EARLIER APPLICATION NUMBER: US 09/062,451
; EARLIER FILING DATE: 1998-04-17
; EARLIER APPLICATION NUMBER: US 08/991,789
; EARLIER FILING DATE: 1997-12-11
; EARLIER APPLICATION NUMBER: US 08/838,762
; EARLIER FILING DATE: 1997-04-09
; EARLIER APPLICATION NUMBER: PCT/US97/00485
; EARLIER FILING DATE: 1997-01-10
; EARLIER APPLICATION NUMBER: US 08/700,014
; EARLIER FILING DATE: 1996-08-20
; EARLIER APPLICATION NUMBER: US 08/585,392
; EARLIER FILING DATE: 1996-01-01
; NUMBER OF SEQ ID NOS: 312
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 60
; LENGTH: 171
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-289-198-60

Query Match 54.8%; Score 18.2; DB 3; Length 171;
Best Local Similarity 52.6%; Pred. No. 1.3e+02;
Matches 20; Conservative 5; Mismatches 13; Indels 0; Gaps 0;

QY 5 AYCNTGGCCNTCCNTAARGAYTAYGARTNAAR 42
Db 134 ATGTGGCATTGACCATTAATACTCTAAGAGGTTAAA 97

RESULT 50
US-09-429-755-60/c
; Sequence 60, Application US/09429755A
```

; Patent No. 6656480  
; GENERAL INFORMATION:  
; APPLICANT: Fridakis, Tony N.  
; APPLICANT: Smith, John M.  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Mishner, Lynda  
; APPLICANT: Retter, Marc W.  
; APPLICANT: Dillon, Davin C.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE  
; TREATMENT AND DIAGNOSIS OF BREAST CANCER  
; FILE REFERENCE: 210121.419C6  
; CURRENT APPLICATION NUMBER: US/09/429,755A  
; CURRENT FILING DATE: 1999-10-28  
; NUMBER OF SEQ ID NOS: 315  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 60  
; LENGTH: 171  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-09-429-755-60

Query Match 54.8%; Score 18.2; DB 3; Length 171;  
Best Local Similarity 52.6%; Pred. No. 1.3e+02;  
Matches 20; Conservative 5; Mismatches 13; Indels 0; Gaps 0;

QY 5 AYCNTGGCCNTCCNTAYAAAGAYTAYGARGTNAAR 42  
Db 134 ATGTGGCATTGACCATTAATACTCTAAGAGGTAA 97

RESULT 51  
US-09-699-295-60/c  
; Sequence 60, Application US/09699295  
; Patent No. 6828431  
; GENERAL INFORMATION:  
; APPLICANT: Fridakis, Tony N.  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Smith, John M.  
; APPLICANT: Mishner, Linda E.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Retter, Marc W.  
; APPLICANT: Wang, Aijun  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Harlocker, Susan L.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE  
; TREATMENT AND DIAGNOSIS OF BREAST CANCER  
; FILE REFERENCE: 210121.419C10  
; CURRENT APPLICATION NUMBER: US/09/699,295  
; CURRENT FILING DATE: 2000-10-26  
; NUMBER OF SEQ ID NOS: 326  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 60  
; LENGTH: 171  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-09-699-295-60

Query Match 54.8%; Score 18.2; DB 3; Length 171;  
Best Local Similarity 52.6%; Pred. No. 1.3e+02;  
Matches 20; Conservative 5; Mismatches 13; Indels 0; Gaps 0;

QY 5 AYCNTGGCCNTCCNTAYAAAGAYTAYGARGTNAAR 42  
Db 134 ATGTGGCATTGACCATTAATACTCTAAGAGGTAA 97

RESULT 52  
US-09-534-825A-60/c  
; Sequence 60, Application US/09534825A  
; Patent No. 6861506  
; GENERAL INFORMATION:  
; APPLICANT: Fridakis, Tony N.  
; APPLICANT: Smith, John M.

; APPLICANT: Reed, Steven G.  
; APPLICANT: Mishner, Lynda  
; APPLICANT: Retter, Marc W.  
; APPLICANT: Dillon, Davin C.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE  
; TREATMENT AND DIAGNOSIS OF BREAST CANCER  
; FILE REFERENCE: 210121.419C7  
; CURRENT APPLICATION NUMBER: US/09/534,825A  
; CURRENT FILING DATE: 2000-03-23  
; NUMBER OF SEQ ID NOS: 317  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 60  
; LENGTH: 171  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-09-534-825A-60

Query Match 54.8%; Score 18.2; DB 3; Length 171;  
Best Local Similarity 52.6%; Pred. No. 1.3e+02;  
Matches 20; Conservative 5; Mismatches 13; Indels 0; Gaps 0;

QY 5 AYCNTGGCCNTCCNTAYAAAGAYTAYGARGTNAAR 42  
Db 134 ATGTGGCATTGACCATTAATACTCTAAGAGGTAA 97

RESULT 53  
US-09-949-016-107890  
; Sequence 107890, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 107890  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-107890

Query Match 54.8%; Score 18.2; DB 3; Length 601;  
Best Local Similarity 48.8%; Pred. No. 2e+02;  
Matches 20; Conservative 6; Mismatches 15; Indels 0; Gaps 0;

QY 2 CNCAYGCGCCNTCCNTAYAAAGAYTAYGARGTNAAR 42  
Db 100 CACATTCATCCCATCCCTTAGAAAAACAATTACTAAAG 140

RESULT 54  
US-09-902-540-1562  
; Sequence 1562, Application US/09902540  
; Patent No. 6833447  
; GENERAL INFORMATION:  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Wiegand, Roger C.  
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(15849)B  
; CURRENT APPLICATION NUMBER: US/09/902,540  
; CURRENT FILING DATE: 2001-07-10

```
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1562
; LENGTH: 669
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-1562

Query Match          54.8%; Score 18.2; DB 3; Length 669;
Best Local Similarity 48.8%; Pred. No. 2e+02;
Matches 20; Conservative 6; Mismatches 15; Indels 0; Gaps 0;

QY 2 CMCAYGCNTGGCCNTCNCNTAYAAARGAYTAYGARGTNAAR 42
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 305 CACATGATTGGGCTTCTAAATGTAAAGATAAGACAAA 345

RESULT 55
US-09-328-352-1269
; Sequence 1269, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: GTC99-03PA
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 1269
; LENGTH: 864
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-1269

Query Match          54.8%; Score 18.2; DB 3; Length 864;
Best Local Similarity 48.8%; Pred. No. 2.2e+02;
Matches 20; Conservative 6; Mismatches 15; Indels 0; Gaps 0;

QY 1 GCNCAYGCNTGGCCNTCNCNTAYAAARGAYTAYGARGTNA 41
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 772 GCCATGATTGGGTGCTACTGTATGAACGCAAGTGAA 812

RESULT 56
US-09-328-352-1336/c
; Sequence 1336, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: GTC99-03PA
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 1336
; LENGTH: 1101
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-1336

Query Match          54.8%; Score 18.2; DB 3; Length 1101;
Best Local Similarity 58.6%; Pred. No. 2.4e+02;
Matches 17; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 13 CCNTCNCNTAYAAARGAYTAYGARGTNA 41
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 676 CCGTCTCGATATCAATATATGAAGTGAA 648

RESULT 57
```

```
US-09-134-000C-2009/c
; Sequence 2009, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE REFERENCE: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: 032796-032
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 2009
; LENGTH: 1335
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-09-134-000C-2009

Query Match          54.8%; Score 18.2; DB 3; Length 1335;
Best Local Similarity 58.6%; Pred. No. 2.5e+02;
Matches 17; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 13 CCNTCNCNTAYAAARGAYTAYGARGTNA 41
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 887 CCTCGCCAAATAAGGATAATGAAGATA 859

RESULT 58
US-09-949-016-15870
; Sequence 15870, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15870
; LENGTH: 45418
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15870

Query Match          54.8%; Score 18.2; DB 3; Length 45418;
Best Local Similarity 48.8%; Pred. No. 7.4e+02;
Matches 20; Conservative 6; Mismatches 15; Indels 0; Gaps 0;

QY 1 GCNCAYGCNTGGCCNTCNCNTAYAAARGAYTAYGARGTNA 41
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 24498 GCTCATGCTTCGTAACTCTCTGTGAATATATAAGGAAA 24538

RESULT 59
US-09-949-016-15494/c
; Sequence 15494, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
```



```
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15494
; LENGTH: 183770
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(183770)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15494

Query Match          54.8%; Score 18.2; DB 3; Length 183770;
Best Local Similarity 54.8%; Pred. No. 1.1e+03;
Matches 17; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 12 GCNCGCNCNTAYAAAGAYTAYGARGTNAAR 42
Db 88582 GCCCTCCATCTACCAAGATTATGAGGGGAAG 88552

RESULT 60
US-09-949-016-15868/c
; Sequence 15868, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15868
; LENGTH: 374159
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15868

Query Match          54.8%; Score 18.2; DB 3; Length 374159;
Best Local Similarity 48.8%; Pred. No. 1.2e+03;
Matches 20; Conservative 6; Mismatches 15; Indels 0; Gaps 0;

QY 1 GCNCGCNCNTGCCNCCNTAYAAAGAYTAYGARGTNA 41
Db 162195 GCCCACTATTACCTTCTCCTTATAAATGTGATGCAGAAAA 162155

RESULT 61
US-09-949-016-14720/c
; Sequence 14720, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
```

```
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14720
; LENGTH: 390890
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(390890)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14720

Query Match          54.8%; Score 18.2; DB 3; Length 390890;
Best Local Similarity 48.8%; Pred. No. 1.3e+03;
Matches 20; Conservative 6; Mismatches 15; Indels 0; Gaps 0;

QY 2 CNCAYGCNTGGCCNCCNTAYAAAGAYTAYGARGTNAAR 42
Db 226842 CACATTTCATCCCATCCCTTAGAAAAACAATTACTATAAG 226802

RESULT 62
US-09-533-559-1061/c
; Sequence 1061, Application US/09533559
; Patent No. 6902887
; GENERAL INFORMATION:
; APPLICANT: Randy M. Berka
; APPLICANT: Michael W. Rey
; APPLICANT: Jeffrey R. Shuster
; APPLICANT: Sakari Kauppinen
; APPLICANT: Ib Groth Clausen
; APPLICANT: Peter Bjørke Olsen
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE OF INVENTION: Expression
; FILE REFERENCE: 5849.200-US
; CURRENT APPLICATION NUMBER: US/09/533,559
; CURRENT FILING DATE: 2000-03-22
; EARLIER APPLICATION NUMBER: 09/273,623
; EARLIER FILING DATE: 1999-03-22
; NUMBER OF SEQ ID NOS: 7860
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1061
; LENGTH: 541
; TYPE: DNA
; ORGANISM: Fusarium venenatum
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(541)
; OTHER INFORMATION: n = A,T,C or G
US-09-533-559-1061

Query Match          54.2%; Score 18; DB 3; Length 541;
Best Local Similarity 55.9%; Pred. No. 2.4e+02;
Matches 19; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

QY 1 GCNCGCNCNTGGCCNCCNCCNTAYAAAGAYTAYG 34
Db 116 GCTCAGCATGCCCCCNCCNCCNCTGTCGCAAGGACTTTG 83

RESULT 63
US-09-949-016-151014
; Sequence 151014, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
```

```

; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: NJ
; COUNTRY: USA
; ZIP: 07068-1739
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,667
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferraro, Gregory D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 325800-435
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 878 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..695
; US-08-469-667-8

Query Match 54.2%; Score 18; DB 2; Length 878;
Best Local Similarity 52.9%; Pred. No. 2.7e+02;
Matches 18; Conservative 6; Mismatches 10; Indels

QY 4 CAYGNTGGCCNTCCNTAYAAAGYATYAGRG 37
   ||:|||||:|||||:|||||:|||||:
Db 710 CATCTGGCATTCACATTTAAAAATATGTGG 743

RESULT 66
US-09-224-110-8
; Sequence 8, Application US/09224110
; Patent No. 6337195
; GENERAL INFORMATION:
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Rosen, Craig
; TITLE OF INVENTION: Colon Specific Genes and Proteins
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carella, Byrne, Bain, Gilfillan, Cecchi,
; ADDRESSEE: Stewart & Olstein
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: NJ
; COUNTRY: USA
; ZIP: 07068-1739
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/224,110
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/469,667
; FILING DATE: 06-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferraro, Gregory D.

```

REGISTRATION NUMBER: 36,134  
REFERENCE/DOCKET NUMBER: 325800-435  
TELEPHONE: 201-994-1700  
TELEFAX: 201-994-1744  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 878 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cdna  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2..685  
US-09-224-110-8

Query Match 54.2%; Score 18; DB 3; Length 878;  
Best Local Similarity 52.9%; Pred. No. 2.7e+02;  
Matches 18; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

Qy 4 CAYCMTGGCCNTCCNTAYARGAYTAYGARG 37  
Db 710 CATTCCTGGCATTACACATTTTAAAAATTATGTGG 743

RESULT 67  
US-09-988-292A-8  
Sequence 8, Application US/09988292A  
Patent No. 6831152  
GENERAL INFORMATION:  
APPLICANT: Yu, Guo-Liang  
TITLE OF INVENTION: Colon Specific Genes and Proteins  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Carella, Byrne, Bain, Gilfillan, Cecchi,  
Stewart & Olstein  
STREET: 6 Becker Farm Road  
CITY: Roseland  
STATE: NJ  
COUNTRY: USA  
ZIP: 07068-1739  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/988,292A  
FILING DATE: 19-Nov. 6831152-2001  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Ferraro, Gregory D.  
REGISTRATION NUMBER: 36,134  
REFERENCE/DOCKET NUMBER: 325800-435  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-994-1700  
TELEFAX: 201-994-1744  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 878 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cdna  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2..685  
SEQUENCE DESCRIPTION: SEQ ID NO: 8:  
US-09-988-292A-8  
Query Match 54.2%; Score 18; DB 3; Length 878;

Best Local Similarity 52.9%; Pred. No. 2.7e+02;  
Matches 18; Conservative 6; Mismatches 10; Indels 0; Gaps 0;  
Qy 4 CAYCMTGGCCNTCCNTAYARGAYTAYGARG 37  
Db 710 CATTCCTGGCATTACACATTTTAAAAATTATGTGG 743

RESULT 68  
PCT-US95-07289-8  
Sequence 8, Application PC/TUS9507289  
GENERAL INFORMATION:  
APPLICANT: Yu, Guo-Liang  
APPLICANT: Rosen, Craig  
TITLE OF INVENTION: Colon Specific Genes and Proteins  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Carella, Byrne, Bain, Gilfillan, Cecchi,  
ADDRESSEE: Stewart & Olstein  
STREET: 6 Becker Farm Road  
CITY: Roseland  
STATE: NJ  
COUNTRY: USA  
ZIP: 07068-1739  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/07289  
FILING DATE: 06-JUN-1995  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Ferraro, Gregory D.  
REGISTRATION NUMBER: 36,134  
REFERENCE/DOCKET NUMBER: 325800-265  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-994-1700  
TELEFAX: 201-994-1744  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 878 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cdna  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2..685  
PCT-US95-07289-8  
Query Match 54.2%; Score 18; DB 7; Length 878;  
Best Local Similarity 52.9%; Pred. No. 2.7e+02;  
Matches 18; Conservative 6; Mismatches 10; Indels 0; Gaps 0;  
Qy 4 CAYCMTGGCCNTCCNTAYARGAYTAYGARG 37  
Db 710 CATTCCTGGCATTACACATTTTAAAAATTATGTGG 743

RESULT 69  
US-09-016-434-850  
Sequence 850, Application US/09016434  
Patent No. 6500938  
GENERAL INFORMATION:  
APPLICANT: Janice Au-Young  
APPLICANT: Jeffrey J. Seilhamer  
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING  
TITLE OF INVENTION: PATHWAY GENE EXPRESSION  
NUMBER OF SEQUENCES: 1490  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.

```

; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016,434
; FILING DATE: HEREWITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0002 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 850:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1512 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: COLNOT01
; CLONE: 608819
; US-09-016-434-850

Query Match 54.2%; Score 18; DB 3; Length 1512;
Best Local Similarity 52.9%; Pred. No. 3.2e+02;
Matches 18; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY 4 CAYGNTGCCNTCCNTAYAAAGAYTAYGARG 37
DB 1400 CATTCCTGGCAITCACATTTTAAAAATTATGTGG 1433

RESULT 70
US-10-044-205A-3
; Sequence 3, Application US/10044205A
; Patent No. 6984502
; GENERAL INFORMATION:
; APPLICANT: KAPELLER-LIBERMANN, Rosana
; APPLICANT: BANDARU, Rajasekhar
; TITLE OF INVENTION: 69087, 15821, and 15418, Methods and Compositions of Human Protein
; TITLE OF INVENTION: Uses Thereof
; FILE REFERENCE: 10147-52U1
; CURRENT APPLICATION NUMBER: US/10/044,205A
; CURRENT FILING DATE: 2002-04-19
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: US 60/242,428
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: US 60/241,884
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: US 60/241,877
; PRIOR FILING DATE: 2000-10-20
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 1659
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-044-205A-3

Query Match 54.2%; Score 18; DB 4; Length 1659;
Best Local Similarity 48.6%; Pred. No. 3.3e+02;
Matches 18; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

US-10-044-205A-3
```

```

Matches 18; Conservative 7; Mismatches 12; Indels 0; Gaps 0;

QY 6 YGCGTGGCCNTCCNTAYAAAGAYTAYGARGTNAAR 42
DB 1155 TGCTGGACGAACACCACTTCAAAAGATTACAAGGAAAAG 1191

RESULT 71
US-09-738-894A-1
; Sequence 1, Application US/09738894A
; Patent No. 6331423
; GENERAL INFORMATION:
; APPLICANT: GUEGLER, Karl et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000636
; CURRENT APPLICATION NUMBER: US/09/738,894A
; CURRENT FILING DATE: 2000-12-18
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1662
; TYPE: DNA
; ORGANISM: Human
; US-09-738-894A-1

Query Match 54.2%; Score 18; DB 3; Length 1662;
Best Local Similarity 48.6%; Pred. No. 3.3e+02;
Matches 18; Conservative 7; Mismatches 12; Indels 0; Gaps 0;

QY 6 YGCGTGGCCNTCCNTAYAAAGAYTAYGARGTNAAR 42
DB 1155 TGCTGGACGAACACCACTTCAAAAGATTACAAGGAAAAG 1191

RESULT 72
US-09-802-117-1
; Sequence 1, Application US/09802117
; Patent No. 6444456
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Wilganowski, Nathaniel L.
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. 644456el Human G-Coupled Protein Receptor Kinases and Polynuc
; TITLE OF INVENTION: Encoding the Same
; FILE REFERENCE: LEX-0147-USA
; CURRENT APPLICATION NUMBER: US/09/802,117
; CURRENT FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: US 60/188,449
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1662
; TYPE: DNA
; ORGANISM: homo sapiens
; US-09-802-117-1

Query Match 54.2%; Score 18; DB 3; Length 1662;
Best Local Similarity 48.6%; Pred. No. 3.3e+02;
Matches 18; Conservative 7; Mismatches 12; Indels 0; Gaps 0;

QY 6 YGCGTGGCCNTCCNTAYAAAGAYTAYGARGTNAAR 42
DB 1155 TGCTGGACGAACACCACTTCAAAAGATTACAAGGAAAAG 1191

RESULT 73
US-09-964-469-1
; Sequence 1, Application US/09964469
; Patent No. 6579709
; GENERAL INFORMATION:
```

```

; APPLICANT: GUEGLER, Karl et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000636DIV
; CURRENT APPLICATION NUMBER: US/09/964,469
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: 60/208,331
; PRIOR FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: 09/738,894
; PRIOR FILING DATE: 2000-12-18
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1662
; TYPE: DNA
; ORGANISM: Human
; US-09-964-469-1

```

```
Query Match      54.2%; Score 18; DB 3; Length 1662;
Best Local Similarity 48.6%; Pred. No. 3.3e+02;
Matches 18: Conservative 7; Mismatches 12; Indels 0; Gaps 0;
```

**Q7**               6 YGNTGGCCNCTNCNCTAAYAARGAYTAYGAGTGTAAR 42  
                :||| || | ||| :|||::||| ::||| :

**D8**             1155 TGCTGGACGAACAACATTCAAGAATTCACAGGAAAG 1191

```

RESULT 74
US-10-217-745-1
; Sequence 1, Application US/10217745
; Patent No. 6838275
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Wilganowski, Nathaniel L.
; APPLICANT: Turnet, C. Alexander Jr.
; TITLE OF INVENTION: No. 6838275el Human G-Coupled Protein Receptor Kinases and
; TITLE OF INVENTION: Polynucleotides
; FILE OF INVENTION: Encoding the Same
; FILE REFERENCE: LEX-0147-USA
; CURRENT APPLICATION NUMBER: US/10/217,745
; CURRENT FILING DATE: 2002-08-12
; PRIOR APPLICATION NUMBER: US/09/802,117
; PRIOR FILING DATE: 2001-03-08
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: fastseq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1662
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-217-745-1

```

```
Query Match      54.2%; Score 18; DB 3; Length 1662;
Best Local Similarity 48.6%; Pred. No. 3.3e+02;
Matches 18: Conservative 7; Mismatches 12; Indels 0; Gaps 0;
```

**Qy** 6 YGNTGGCNCNTAATCAAGAYTAYGARTNAAR 42  
:  
**Db** 1155 TGCTGGAGCAACACCATTCAAGATTACAAGGAAGAAG 1191

```

RESULT 75
US-10-425-962-1
; Sequence 1, Application US/10425962
; Patent No. 7029894
; GENERAL INFORMATION:
; APPLICANT: GUEGLER, Karl et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000636DIV2
; CURRENT APPLICATION NUMBER: US/10/425,962
; CURRENT FILING DATE: 2003-04-30

```

```

; PRIOR APPLICATION NUMBER: 09/964,469
; PRIOR FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: 09/738,894
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: 60/208,331
; PRIOR FILING DATE: 2000-06-01
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: fastseq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1662
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-425-962-1

```

Query Match 54.2%; Score 18; DB 5; Length 1662;  
Best Local Similarity 48.6%; Pred. No. 3.3e+02;  
Matches 18: Conservative 7; Mismatches 12; Indels

**Qy**

6 YGCNTGGCCNCTCNCNTAYAAARGAYTAYGARGTGNAAR 42  
          :|||       |||       |||       |||       |||       |||       |||       |||       |||  
**pB**      1155 TGCTGTGAACGAAACCATTCAAGATTCACAAAGAAAAAG 1191

```

RESULT 76
US-09-583-110-560
, Sequence 560, Application US/09583110
, Patent No. 6699703
, GENERAL INFORMATION:
, APPLICANT: Lyvyn Doucette-Stamm et al
, TITLE OF INVENTION: Nucleic Acid an
, TITLE OF INVENTION: Pneumoniae for
, FILE REFERENCE: PATH00-07A
, CURRENT APPLICATION NUMBER: US/09/58
, CURRENT FILING DATE: 2000-05-26
, PRIOR APPLICATION NUMBER: US 09/107
, PRIOR FILING DATE: 1998-06-30
, PRIOR APPLICATION NUMBER: US 60/085
, PRIOR FILING DATE: 1998-05-12
, PRIOR APPLICATION NUMBER: US 60/051
, PRIOR FILING DATE: 1997-07-02
, NUMBER OF SEQ ID NOS: 5322
, SEQ ID NO 560
, LENGTH: 1953
, TYPE: DNA
, ORGANISM: Streptococcus pneumoniae
US-09-583-110-560

```

Query Match 54.2%; Score 18; DB 3; Length 1953;  
Best Local Similarity 52.5%; Pred. No. 3.5e+02;  
Matches 21: Conservative 4; Mismatches 15; Indels 0; Gaps 0;

QY 2 CNCA YGCNTGGCCNTCNCNTAY AARGAYTAYGARGTNA 41

Dp 1208 CAGAAGAA TTGCCTGTCACGTTACAAGGATAAGGTAGTGAA 1247

```

RESULT 77
US-09-107-433-912
; Sequence 912, Application US/09107433
; Patent No. 6800744
; GENERAL INVENTION:
; APPLICANT: LYNN A DOUCETTE-STRA
; TITLE OF INVENTION: NUCLEIC AC
; SEQUENCES
; THERAPEUTIC
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPE
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354

```

FOR DIAGNO

COMPUTER READABLE FORM:  
MEDIUM TYPE: CD/ROM ISO9660  
COMPUTER: <Unknown>  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: <Unknown>  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,433  
FILING DATE: 30-Jun-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/ 085131  
FILING DATE: May 12, 1998  
APPLICATION NUMBER: 60/051553  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Ariniello, Pamela Deneka  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-011  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 912:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1953 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Streptococcus pneumoniae  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: (B) LOCATION 1...1953  
SEQUENCE DESCRIPTION: SEQ ID NO: 912:

US-09-107-433-912

Query Match 54.2%; Score 18; DB 3; Length 1953;  
Best Local Similarity 52.5%; Pred. No. 3.5e+02;  
Matches 21; Conservative 4; Mismatches 15; Indels 0; Gaps 0;

QY 2 CNCAYCGTCGTCNTCNCNTAARGAYTAYGARGTNA 41  
DB 1208 CAGAGAAATTCGTCGACGTTACAGGATAAGTAGTGAA 1247

RESULT 78  
US-09-248-796A-837  
Sequence 837, Application US/09248796A  
Patent No. 6747137  
GENERAL INFORMATION:  
APPLICANT: Keith Weinstein et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS  
TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.132  
CURRENT APPLICATION NUMBER: US/09/248,796A  
CURRENT FILING DATE: 1999-02-12  
PRIOR APPLICATION NUMBER: US 60/074,725  
PRIOR FILING DATE: 1998-02-13  
PRIOR APPLICATION NUMBER: US 60/096,409  
PRIOR FILING DATE: 1998-08-13  
NUMBER OF SEQ ID NOS: 28208  
SEQ ID NO 837  
LENGTH: 2148  
TYPE: DNA  
ORGANISM: Candida albicans  
US-09-248-796A-837

Query Match 54.2%; Score 18; DB 3; Length 2148;  
Best Local Similarity 52.5%; Pred. No. 3.6e+02;  
Matches 21; Conservative 4; Mismatches 15; Indels 0; Gaps 0;

QY 1 GCNCAYCGTCGTCNTCNCNTAARGAYTAYGARGTNA 40

Db 769 GCCAGGCATCGCAATTTCTTTGAAAGATGTCGAGGTCA 808

RESULT 79  
US-10-044-205A-1  
Sequence 1, Application US/10044205A  
Patent No. 6984502  
GENERAL INFORMATION:  
APPLICANT: KAPPELLER-LIBERMANN, Rosana  
APPLICANT: BANDARU, Rajasekhar  
TITLE OF INVENTION: 69087, 15821, and 15418, Methods and Compositions of Human Protein  
TITLE OF INVENTION: Uses Thereof  
FILE REFERENCE: 10147-52U1  
CURRENT APPLICATION NUMBER: US/10/044,205A  
CURRENT FILING DATE: 2002-04-19  
PRIOR APPLICATION NUMBER: US 60/242,428  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: US 60/241,884  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: US 60/241,877  
PRIOR FILING DATE: 2000-10-20  
NUMBER OF SEQ ID NOS: 44  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 1  
LENGTH: 2198  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-044-205A-1

Query Match 54.2%; Score 18; DB 4; Length 2198;  
Best Local Similarity 48.6%; Pred. No. 3.6e+02;  
Matches 18; Conservative 7; Mismatches 12; Indels 0; Gaps 0;

QY 6 YGCNTGGCCNTCNCNTAARGAYTAYGARGTNAAR 42  
DB 1445 TGCTGACGACACCACTTCAAGATTACAGGAAAAG 1481

RESULT 80  
US-09-802-117-5  
Sequence 5, Application US/09802117  
Patent No. 6444456  
GENERAL INFORMATION:  
APPLICANT: Walke, D. Wade  
APPLICANT: Wilganowski, Nathaniel L.  
APPLICANT: Turner, C. Alexander Jr.  
TITLE OF INVENTION: No. 6444456el Human G-Coupled Protein Receptor Kinases and Polymers  
TITLE OF INVENTION: Encoding the Same  
FILE REFERENCE: LEX-0147-USA  
CURRENT APPLICATION NUMBER: US/09/802,117  
CURRENT FILING DATE: 2001-03-08  
PRIOR APPLICATION NUMBER: US 60/188,449  
PRIOR FILING DATE: 2000-03-10  
NUMBER OF SEQ ID NOS: 5  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 5  
LENGTH: 2249  
TYPE: DNA  
ORGANISM: homo sapiens  
US-09-802-117-5

Query Match 54.2%; Score 18; DB 3; Length 2249;  
Best Local Similarity 48.6%; Pred. No. 3.7e+02;  
Matches 18; Conservative 7; Mismatches 12; Indels 0; Gaps 0;

QY 6 YGCNTGGCCNTCNCNTAARGAYTAYGARGTNAAR 42  
DB 1508 TGCTGACGACACCACTTCAAGATTACAGGAAAAG 1544

RESULT 81  
US-10-217-745-5

; Sequence 5, Application US/10217745  
; Patent No. 6838275  
; GENERAL INFORMATION:  
; APPLICANT: Walke, D. Wade  
; APPLICANT: Wilganowski, Nathaniel L.  
; APPLICANT: Turner, C. Alexander Jr.  
; TITLE OF INVENTION: No. 6838275el Human G-Coupled Protein Receptor Kinases and  
; TITLE OF INVENTION: Polynucleotides  
; TITLE OF INVENTION: Encoding the Same  
; FILE REFERENCE: LEX-0147-USA  
; CURRENT APPLICATION NUMBER: US/10/217,745  
; CURRENT FILING DATE: 2002-08-12  
; PRIOR APPLICATION NUMBER: US/09/802,117  
; PRIOR FILING DATE: 2001-03-08  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 2249  
; TYPE: DNA  
; ORGANISM: homo sapiens  
US-10-217-745-5

Query Match 54.2%; Score 18; DB 3; Length 2249;  
Best Local Similarity 48.6%; Pred. No. 3.7e+02;  
Matches 18; Conservative 7; Mismatches 12; Indels 0; Gaps 0;

Qy 6 YGNTGGCCNTCCNTAYAARGAYTAYGARGTNAAR 42  
Db 1508 TGCTGGACGAACCACTTCAAGATTACAGGAAAG 1544

RESULT 82  
US-08-757-541-4/c  
; Sequence 4, Application US/08757541  
; Patent No. 5766877  
; GENERAL INFORMATION:  
; APPLICANT: Stark, Kevin Lee  
; APPLICANT: Luethy, Roland  
; TITLE OF INVENTION: NOVEL AGOUTI-RELATED GENE  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: AMGEN INC.  
; STREET: 1840 DEHAVILLAND DRIVE  
; CITY: THOUSAND OAKS  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 91320-1789  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/757,541  
; FILING DATE:  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: OLESKI, NANCY A  
; REGISTRATION NUMBER: 34,688  
; REFERENCE/DOCKET NUMBER: A-402A  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2371 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-757-541-4

Query Match 54.2%; Score 18; DB 2; Length 2371;  
Best Local Similarity 50.0%; Pred. No. 3.7e+02;  
Matches 21; Conservative 5; Mismatches 16; Indels 0; Gaps 0;

Qy 1 GCNCAYGCNTGGCCNTCCNTAYAARGAYTAYGARGTNAAR 42  
Db 1728 GCACACTCTCTGGCTTCCCTCTGGTAACAGGATGAGGGCAG 1687

RESULT 83  
US-09-033-275-4/c  
; Sequence 4, Application US/09033275  
; Patent No. 6060589  
; GENERAL INFORMATION:  
; APPLICANT: Stark, Kevin Lee  
; APPLICANT: Luethy, Roland  
; TITLE OF INVENTION: NOVEL AGOUTI-RELATED GENE  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: AMGEN INC.  
; STREET: 1840 DEHAVILLAND DRIVE  
; CITY: THOUSAND OAKS  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 91320-1789  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/033,275  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/757,541  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: OLESKI, NANCY A  
; REGISTRATION NUMBER: 34,688  
; REFERENCE/DOCKET NUMBER: A-402A  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2371 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-09-033-275-4

Query Match 54.2%; Score 18; DB 3; Length 2371;  
Best Local Similarity 50.0%; Pred. No. 3.7e+02;  
Matches 21; Conservative 5; Mismatches 16; Indels 0; Gaps 0;

Qy 1 GCNCAYGCNTGGCCNTCCNTAYAARGAYTAYGARGTNAAR 42  
Db 1728 GCACACTCTCTGGCTTCCCTCTGGTAACAGGATGAGGGCAG 1687

RESULT 84  
US-09-342-581-4/c  
; Sequence 4, Application US/09342581  
; Patent No. 6203995  
; GENERAL INFORMATION:  
; APPLICANT: Stark, Kevin Lee  
; APPLICANT: Luethy, Roland  
; TITLE OF INVENTION: NOVEL AGOUTI-RELATED GENE  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: AMGEN INC.  
; STREET: 1840 DEHAVILLAND DRIVE  
; CITY: THOUSAND OAKS  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 91320-1789  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk





; APPLICANT: Beetham, P.  
; APPLICANT: Avissar, P.  
; APPLICANT: Walker, K.  
; APPLICANT: Metz, R.

; TITLE OF INVENTION: NON-TRANSGENIC HERBICIDE RESISTANT PLANTS

; FILE REFERENCE: 7991-086

; CURRENT APPLICATION NUMBER: US/09/685,403

; CURRENT FILING DATE: 2000-10-10

; PRIOR APPLICATION NUMBER: 60/158,027

; PRIOR FILING DATE: 1999-10-07

; PRIOR APPLICATION NUMBER: 60/173,564

; PRIOR FILING DATE: 1999-12-30

; NUMBER OF SEQ ID NOS: 44

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 1

; LENGTH: 2763

; TYPE: DNA

; ORGANISM: Arabidopsis thaliana

US-09-685-403-1

Query Match 54.2%; Score 18; DB 3; Length 2763;

Best Local Similarity 52.5%; Pred. No. 3.9e+02;

Matches 21; Conservative 4; Mismatches 15; Indels 0; Gaps 0;

Qy 2 CNCAYGCGTCCNTCCNTCCNTAYAARGAYTAYGARGTNA 41

Db 1771 CACATCCCTTAGCTTGACATATATGACTAAAGGTGAA 1810

RESULT 88

US-09-193-562D-27

; Sequence 27, Application US/09193562D

; Patent No. 6309857

; GENERAL INFORMATION:

; APPLICANT: Pauli, Benedicht U.

; TITLE OF INVENTION: Nucleotide Sequences Encoding Mammalian Calcium

; FILE REFERENCE: 18617,0052

; CURRENT APPLICATION NUMBER: US/09/193,562D

; CURRENT FILING DATE: 1998-11-17

; PRIOR APPLICATION NUMBER: US/60/065,922

; PRIOR FILING DATE: 1997-11-17

; NUMBER OF SEQ ID NOS: 47

; SEQ ID NO 27

; LENGTH: 3007

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-193-562D-27

Query Match 54.2%; Score 18; DB 3; Length 3007;

Best Local Similarity 52.9%; Pred. No. 4e+02;

Matches 18; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

Qy 4 CAYGCGTCCNTCCNTCCNTAYAARGAYTAYGARG 37

Db 2722 CATTCCTGGCATTACACATTTTAAAAATTAATGTGG 2755

RESULT 89

US-10-055-412B-27

; Sequence 27, Application US/10055412B

; Patent No. 6692939

; GENERAL INFORMATION:

; APPLICANT: Pauli, Benedicht U.

; TITLE OF INVENTION: Nucleotide Sequences Encoding Mammalian Calcium

; FILE REFERENCE: 18617,0058

; CURRENT APPLICATION NUMBER: US/10/055,412B

; CURRENT FILING DATE: 2001-10-29

; PRIOR APPLICATION NUMBER: US/09/193,562

; PRIOR FILING DATE: 1998-11-17

; PRIOR APPLICATION NUMBER: US/60/065,922

; PRIOR FILING DATE: 1997-11-17

; NUMBER OF SEQ ID NOS: 47

; SEQ ID NO 27

; LENGTH: 3007

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-055-412B-27

Query Match 54.2%; Score 18; DB 3; Length 3007;

Best Local Similarity 52.9%; Pred. No. 4e+02;

Matches 18; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

Qy 4 CAYGCGTCCNTCCNTCCNTAYAARGAYTAYGARG 37

Db 2722 CATTCCTGGCATTACACATTTTAAAAATTAATGTGG 2755

RESULT 90

US-08-961-527-120/c

; Sequence 120, Application US/08961527

; Patent No. 6420135

; GENERAL INFORMATION:

; APPLICANT: Charles Kunsch

; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences

; NUMBER OF SEQUENCES: 391

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

; COMPUTER: HP Vectra 486/33

; OPERATING SYSTEM: MSDOS version 6.2

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/961,527

; FILING DATE:

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Brookes, A. Anders

; REGISTRATION NUMBER: 36,373

; REFERENCE/DOCKET NUMBER: PB340P1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (301) 309-8504

; TELEFAX: (301) 309-8512

; INFORMATION FOR SEQ ID NO: 120:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 7172 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

US-08-961-527-120

Query Match 54.2%; Score 18; DB 3; Length 7172;

Best Local Similarity 52.5%; Pred. No. 5.2e+02;

Matches 21; Conservative 4; Mismatches 15; Indels 0; Gaps 0;

Qy 2 CNCAYGCGTCCNTCCNTCCNTAYAARGAYTAYGARGTNA 41

Db 5978 CAGAAGAAATTCCTGCACGTACAAAGGATAAGTAGTGAA 5939

RESULT 91

US-10-001-189-67

; Sequence 67, Application US/10001189

; Patent No. 6962810

; GENERAL INFORMATION:

; APPLICANT: FRASER JR., MALCOLM J.

```
; APPLICANT: LI, XU
; APPLICANT: BEAM, TERESA
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TRANSPOSITION USING
; TITLE OF INVENTION: MINIMAL SEGMENTS OF THE EUKARYOTIC TRANSFORMATION
; TITLE OF INVENTION: VECTOR PIGGYBAC
; FILE REFERENCE: 835910-92098
; CURRENT APPLICATION NUMBER: US/10/001,189
; CURRENT FILING DATE: 2001-10-30
; PRIOR APPLICATION NUMBER: 60/244,984
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: 60/244,677
; PRIOR FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 70
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 67
; LENGTH: 7411
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: pBSII-Act5c-ori sequence
US-10-001-189-67

Query Match          54.2%; Score 18; DB 3; Length 7411;
Best Local Similarity 51.4%; Pred. No. 5.3e+02;
Matches 18; Conservative 6; Mismatches 11; Indels 0; Gaps 0;

QY      8  CMTGGCCNTCNCNTAYAAAGAYTAYGARGTNAAR 42
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      2849  CGTGACCATCACAGCATAAAGATACCGCTGAAG 2883

RESULT 92
US-09-949-016-17104
; Sequence 17104, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17104
; LENGTH: 15566
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-17104

Query Match          54.2%; Score 18; DB 3; Length 15566;
Best Local Similarity 56.8%; Pred. No. 6.7e+02;
Matches 21; Conservative 3; Mismatches 13; Indels 0; Gaps 0;

QY      1  GCNAYGCNTGCCNTCNCNTAYAAAGAYTAYGARG 37
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      1893  GCAGAAGCCTGACCATCCCTAGAGGGCCTGGGAGG 1929

RESULT 93
US-09-738-894A-3
; Sequence 3, Application US/09738894A
; Patent No. 6331423
; GENERAL INFORMATION:
; APPLICANT: GUEGLER, Karl et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
```

```
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000636
; CURRENT APPLICATION NUMBER: US/09/738,894A
; CURRENT FILING DATE: 2000-12-18
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 36651
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(36651)
; OTHER INFORMATION: n = A,T,C or G
US-09-738-894A-3

Query Match          54.2%; Score 18; DB 3; Length 36651;
Best Local Similarity 48.6%; Pred. No. 8.6e+02;
Matches 18; Conservative 7; Mismatches 12; Indels 0; Gaps 0;

QY      6  YGCNTGGCCNTCNCNTAYAAAGAYTAYGARGTNAAR 42
      :||| ||||| ||||| ||||| ||||| ||||| |||||
Db      31336  TGCTGGACGACACACCATTCAAAGATTACAAGGAAAAG 31372

RESULT 94
US-09-964-469-3
; Sequence 3, Application US/09964469
; Patent No. 6579709
; GENERAL INFORMATION:
; APPLICANT: GUEGLER, Karl et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; FILE REFERENCE: CL000636DIV
; CURRENT APPLICATION NUMBER: US/09/964,469
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: 60/208,331
; PRIOR FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: 09/738,894
; PRIOR FILING DATE: 2000-12-18
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 36651
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(36651)
; OTHER INFORMATION: n = A,T,C or G
US-09-964-469-3

Query Match          54.2%; Score 18; DB 3; Length 36651;
Best Local Similarity 48.6%; Pred. No. 8.6e+02;
Matches 18; Conservative 7; Mismatches 12; Indels 0; Gaps 0;

QY      6  YGCNTGGCCNTCNCNTAYAAAGAYTAYGARGTNAAR 42
      :||| ||||| ||||| ||||| ||||| ||||| |||||
Db      31336  TGCTGGACGACACACCATTCAAAGATTACAAGGAAAAG 31372

RESULT 95
US-10-425-962-3
; Sequence 3, Application US/10425962
; Patent No. 7029894
; GENERAL INFORMATION:
; APPLICANT: GUEGLER, Karl et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000636DIV2
```

10 ; CURRENT AFFILIATION NUMBER: 05/05/2437,

; PRIOR FILING DATE: 2000-10-03

**CURRENT AFFAIRS**



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2006, 15:34:22 ; Search time 20.3 Seconds  
(without alignments)  
262.600 Million cell updates/sec

Title: US-10-600-816-33  
Perfect score: 33.2  
Sequence: 1 gncaygctggcctncc.....ayaargaytaygargtarnaar 42

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 253354 seqs, 63461778 residues

Total number of hits satisfying chosen parameters: 506708

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 1000 summaries

Database : Published Applications\_NA\_New.\*  
1: /EMC\_Celerra\_SIDS3/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*  
2: /EMC\_Celerra\_SIDS3/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*  
3: /EMC\_Celerra\_SIDS3/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*  
4: /EMC\_Celerra\_SIDS3/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*  
5: /EMC\_Celerra\_SIDS3/ptodata/1/pubpna/FCT\_NEW\_PUB.seq.\*  
6: /EMC\_Celerra\_SIDS3/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*  
7: /EMC\_Celerra\_SIDS3/ptodata/1/pubpna/US11\_NEW\_PUB.seq.\*  
8: /EMC\_Celerra\_SIDS3/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	63.3	1101	7	US-11-217-529-82765 Sequence 82765, A
2	20.4	61.4	2292	7	US-11-217-529-78725 Sequence 78725, A
3	18.8	56.6	1161	7	US-11-217-529-5462 Sequence 5462, A
4	18.8	56.6	2871	7	US-11-217-529-76779 Sequence 76779, A
5	18.4	55.4	1624	6	US-10-953-349-1004 Sequence 1004, A
6	18.2	54.8	2118	7	US-11-217-529-75440 Sequence 75440, A
7	18	54.2	118899	7	US-11-189-279-64 Sequence 64, Appl
8	17.8	53.6	2676	7	US-11-217-529-5102 Sequence 5102, Ap
9	17.6	53.0	818	6	US-10-953-349-38583 Sequence 38583, A
10	17.6	53.0	939	7	US-11-217-529-2422 Sequence 2422, Ap
11	17.6	53.0	54550	7	US-11-318-813-42 Sequence 42, Appl
12	17.4	52.4	955	6	US-10-488-619-3026 Sequence 3026, Ap
13	17.4	52.4	1269	6	US-10-471-571A-5437 Sequence 5437, Ap
14	17.4	52.4	1674	7	US-11-217-529-76209 Sequence 76209, A
15	17.4	52.4	1731	7	US-11-217-529-3913 Sequence 3913, Ap
16	17.4	52.4	1897	6	US-10-505-928-269 Sequence 269, App
17	17.4	52.4	2876	7	US-11-293-697-1526 Sequence 1526, Ap
18	17.2	51.8	972	6	US-10-471-571A-5233 Sequence 523, App
19	17.2	51.8	1127	6	US-10-953-349-2306 Sequence 2306, A
20	17.2	51.8	1377	7	US-11-217-529-78099 Sequence 78099, A
21	17	51.2	1349	6	US-10-953-349-29678 Sequence 29678, A
22	17	51.2	1731	7	US-11-217-529-754 Sequence 754, App
23	17	51.2	1800	7	US-11-217-529-1835 Sequence 1835, Ap
24	17	51.2	2269	7	US-11-293-697-51 Sequence 51, Appl
25	17	51.2	2856	7	US-11-293-697-2015 Sequence 2015, Ap

17	51.2	3007	6	US-10-953-349-31575 Sequence 31575, A
17	51.2	3008	7	US-11-293-697-1736 Sequence 1736, Ap
17	51.2	3743	6	US-10-502-993-1 Sequence 1, Appli
29	51.2	4656	7	US-11-217-529-3130 Sequence 3130, Ap
30	51.2	5003	6	US-10-473-173-112 Sequence 112, App
31	50.6	802	6	US-10-953-349-679 Sequence 679, App
32	50.6	2358	7	US-11-293-697-245 Sequence 245, App
33	50.6	3156	7	US-11-293-697-537 Sequence 537, App
34	50.6	3567	7	US-11-311-778-19 Sequence 19, Appli
35	50.6	5193	7	US-11-270-040-5 Sequence 5, Appli
36	50.6	54550	7	US-11-318-813-42 Sequence 42, Appli
37	50.6	138941	6	US-10-489-730-10 GENERAL INFORMATI
38	50.6	595	6	US-10-488-619-2144 Sequence 2144, Ap
39	50.6	644	6	US-10-953-349-14311 Sequence 14311, A
40	50.6	851	6	US-10-488-619-2143 Sequence 2143, Ap
41	50.6	1266	6	US-10-953-349-31441 Sequence 31441, A
42	50.6	1312	6	US-10-953-349-25421 Sequence 25421, A
43	50.6	1746	6	US-10-953-349-8766 Sequence 8766, Ap
44	50.6	2106	7	US-11-217-529-76864 Sequence 76864, A
45	50.6	3306	7	US-11-217-529-78513 Sequence 78513, A
46	50.6	530	6	US-10-953-349-13812 Sequence 13812, A
47	49.4	698	6	US-10-488-619-1977 Sequence 1977, Ap
48	49.4	909	6	US-10-488-619-1795 Sequence 1795, Ap
49	49.4	1087	6	US-10-953-349-39951 Sequence 39951, A
50	49.4	1228	6	US-10-953-349-24567 Sequence 24567, A
51	49.4	1406	6	US-10-953-349-35818 Sequence 35818, A
52	49.4	1462	6	US-10-953-349-10383 Sequence 10383, A
53	49.4	1566	7	US-11-217-529-79698 Sequence 79698, A
54	49.4	1854	7	US-11-293-697-2093 Sequence 2093, Ap
55	49.4	2180	6	US-10-505-928-771 Sequence 771, App
56	49.4	2235	7	US-11-217-529-76276 Sequence 76276, A
57	49.4	2426	7	US-11-293-697-834 Sequence 834, App
58	49.4	2700	7	US-11-293-697-289 Sequence 289, App
59	49.4	3315	6	US-10-505-928-554 Sequence 554, App
60	48.8	200	7	US-11-329-010-4 Sequence 4, Appli
61	48.8	423	6	US-10-488-619-208 Sequence 208, App
62	48.8	742	7	US-11-301-554-1345 Sequence 1345, Ap
63	48.8	824	6	US-10-953-349-25896 Sequence 25896, A
64	48.8	1365	6	US-10-953-349-21915 Sequence 21915, A
65	48.8	1424	6	US-10-953-349-20753 Sequence 20753, A
66	48.8	1506	7	US-11-217-529-75494 Sequence 75494, A
67	48.8	1579	7	US-11-293-697-1142 Sequence 1142, Ap
68	48.8	1632	7	US-11-217-529-2174 Sequence 2174, Ap
69	48.8	1745	6	US-10-953-349-2142 Sequence 2142, Ap
70	48.8	1828	6	US-10-953-349-9257 Sequence 9257, Ap
71	48.8	1831	6	US-10-953-349-2382 Sequence 2382, Ap
72	48.8	2163	7	US-11-217-529-4479 Sequence 4479, Ap
73	48.8	2163	7	US-11-217-529-6203 Sequence 6203, Ap
74	48.8	2280	7	US-11-217-529-1353 Sequence 1353, Ap
75	48.8	2298	7	US-11-293-697-1375 Sequence 1375, Ap
76	48.8	2413	6	US-10-953-349-40221 Sequence 40221, A
77	48.8	2825	6	US-10-471-571A-4801 Sequence 4801, Ap
78	48.8	5479	6	US-10-505-928-472 Sequence 472, App
79	48.8	7603	6	US-10-511-937-2852 Sequence 2852, Ap
80	48.2	607	7	US-11-301-554-489 Sequence 489, App
81	48.2	735	7	US-11-217-529-79889 Sequence 79889, A
82	48.2	762	6	US-10-471-571A-3165 Sequence 3165, Ap
83	48.2	914	6	US-10-953-349-6898 Sequence 6898, Ap
84	48.2	1227	6	US-10-953-349-35524 Sequence 35524, A
85	48.2	1358	6	US-10-953-349-29958 Sequence 29958, A
86	48.2	1371	6	US-10-953-349-21144 Sequence 21144, A
87	48.2	1499	6	US-10-953-349-24287 Sequence 24287, A
88	48.2	1584	6	US-10-953-349-12909 Sequence 12909, A
89	48.2	1948	1	US-09-949-925-71 Sequence 71, Appl
90	48.2	1967	7	US-11-154-977-2 Sequence 2, Appli
91	48.2	2079	7	US-11-217-529-76446 Sequence 286, App
92	48.2	2254	7	US-11-293-697-286 Sequence 286, App
93	48.2	2301	7	US-11-217-529-2992 Sequence 2992, Ap
94	48.2	2320	7	US-11-293-697-1298 Sequence 1298, Ap
95	48.2	2394	7	US-11-217-529-81691 Sequence 81691, A
96	48.2	2500	7	US-11-293-697-1078 Sequence 1078, Ap
97	48.2	2634	7	US-11-217-529-79043 Sequence 79043, A
98	48.2	2793	7	US-11-217-529-3383 Sequence 3383, Ap















```
c 975 14.2 42.8 768 7 US-11-266-747-2259 Sequence 2259, Ap
c 976 14.2 42.8 768 7 US-11-266-747-2261 Sequence 2261, Ap
c 977 14.2 42.8 768 7 US-11-266-747-2263 Sequence 2263, Ap
c 978 14.2 42.8 768 7 US-11-266-747-2265 Sequence 2265, Ap
c 979 14.2 42.8 768 7 US-11-266-747-2267 Sequence 2267, Ap
c 980 14.2 42.8 768 7 US-11-266-747-2269 Sequence 2269, Ap
c 981 14.2 42.8 768 7 US-11-266-747-2271 Sequence 2271, Ap
c 982 14.2 42.8 768 7 US-11-266-747-2273 Sequence 2273, Ap
c 983 14.2 42.8 768 7 US-11-266-747-2275 Sequence 2275, Ap
c 984 14.2 42.8 768 7 US-11-266-747-2277 Sequence 2277, Ap
c 985 14.2 42.8 768 7 US-11-266-747-2279 Sequence 2279, Ap
c 986 14.2 42.8 768 7 US-11-266-747-2281 Sequence 2281, Ap
c 987 14.2 42.8 768 7 US-11-266-747-2283 Sequence 2283, Ap
c 988 14.2 42.8 768 7 US-11-266-747-2285 Sequence 2285, Ap
c 989 14.2 42.8 768 7 US-11-266-747-2287 Sequence 2287, Ap
c 990 14.2 42.8 768 7 US-11-266-747-2289 Sequence 2289, Ap
c 991 14.2 42.8 768 7 US-11-266-747-2291 Sequence 2291, Ap
c 992 14.2 42.8 768 7 US-11-266-747-2293 Sequence 2293, Ap
c 993 14.2 42.8 768 7 US-11-266-747-2295 Sequence 2295, Ap
c 994 14.2 42.8 768 7 US-11-266-747-2297 Sequence 2297, Ap
c 995 14.2 42.8 768 7 US-11-266-747-2301 Sequence 2301, Ap
c 996 14.2 42.8 768 7 US-11-266-747-2303 Sequence 2303, Ap
c 997 14.2 42.8 768 7 US-11-266-747-2305 Sequence 2305, Ap
c 998 14.2 42.8 768 7 US-11-266-747-2307 Sequence 2307, Ap
c 999 14.2 42.8 768 7 US-11-266-747-2309 Sequence 2309, Ap
c1000 14.2 42.8 768 7 US-11-266-747-2315 Sequence 2315, Ap
```

## ALIGNMENTS

```
RESULT 1
US-11-217-529-82765
; Sequence 82765, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; PRIOR FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: Patentin version 3.3
; SEQ ID NO 82765
; LENGTH: 1101
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-82765
```

```
Query Match 63.3%; Score 21; DB 7; Length 1101;
Best Local Similarity 55.3%; Pred. No. 0.67;
Matches 21; Conservative 6; Mismatches 11; Indels 0; Gaps 0;
```

```
QY 4 CAYGNTGGCCNTCCNTAYARGAYTAYGAGTNA 41
||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 676 CATGCGTGGCCCACTCAACAAAATAATGAAGTTAA 713
```

```
RESULT 2
US-11-217-529-78725
; Sequence 78725, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
```

```
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: Patentin version 3.3
; SEQ ID NO 78725
; LENGTH: 2292
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-78725
```

```
Query Match 61.4%; Score 20.4; DB 7; Length 2292;
Best Local Similarity 56.8%; Pred. No. 1.6;
Matches 21; Conservative 5; Mismatches 11; Indels 0; Gaps 0;
```

```
QY 6 YGNTGGCCNTCCNTAYARGAYTAYGAGTNAAR 42
||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2228 CGCTGGCCATCACCATACATTGGTTAGGAGGAAA 2264
```

```
RESULT 3
US-11-217-529-5462
; Sequence 5462, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: Patentin version 3.3
; SEQ ID NO 5462
; LENGTH: 1161
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-5462
```

```
Query Match 56.6%; Score 18.8; DB 7; Length 1161;
Best Local Similarity 54.1%; Pred. No. 7.3;
Matches 20; Conservative 5; Mismatches 12; Indels 0; Gaps 0;
```

```
QY 5 AYGNTGGCCNTCCNTAYARGAYTAYGAGTNA 41
||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 692 ATGCGTGTCTCATCGAACAAAGAAAACGAAGTCAA 728
```

```
RESULT 4
US-11-217-529-76779
; Sequence 76779, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
```

```
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR FILING DATE: 2004-09-02
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: Patentin version 3.3
; SEQ ID NO 76779
; LENGTH: 2871
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-76779

Query Match      56.6%; Score 18.8; DB 7; Length 2871;
Best Local Similarity 54.1%; Pred. No. 10;
Matches 20; Conservative 5; Mismatches 12; Indels 0; Gaps 0;

QY 6 YGNTGGCCNTCCNTAYAARGAYTAYGARGTNAAR 42
Db 2469 TGCATCCGCTCTCTCCTGAGCAAGATCAAGAAGCAAG 2505

RESULT 5
US-10-953-349-1004
; Sequence 1004, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: Patentin version 3.3
; SEQ ID NO 1004
; LENGTH: 1624
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-10-953-349-1004

Query Match      55.4%; Score 18.4; DB 6; Length 1624;
Best Local Similarity 54.3%; Pred. No. 13;
Matches 19; Conservative 5; Mismatches 11; Indels 0; Gaps 0;

QY 8 CNTGGCCNTCCNTAYAARGAYTAYGARGTNAAR 42
Db 154 CCTGCCATCACCGTACTAAGCATGAAGATAAA 188

RESULT 6
US-11-217-529-75440
; Sequence 75440, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: Patentin version 3.3
; SEQ ID NO 75440
; LENGTH: 2118
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-75440
```

```
Query Match      54.8%; Score 18.2; DB 7; Length 2118;
Best Local Similarity 48.8%; Pred. No. 17;
Matches 20; Conservative 6; Mismatches 15; Indels 0; Gaps 0;

QY 2 CNCAYGCNTGCCNTCCNTAYAARGAYTAYGARGTNAAR 42
Db 304 CCCTGTTGGCTTAGTATTCCAAAGACTCCGACTTTAAA 344

RESULT 7
US-11-189-279-64/c
; Sequence 64, Application US/11189279
; Publication No. US20060115829A1
; GENERAL INFORMATION:
; APPLICANT: MAO, LI
; APPLICANT: WANG, JIE
; APPLICANT: LUO, WANG
; TITLE OF INVENTION: A METHOD OF TREATING CANCER
; FILE REFERENCE: UTXC:875US
; CURRENT APPLICATION NUMBER: US/11/189,279
; CURRENT FILING DATE: 2005-07-26
; PRIOR APPLICATION NUMBER: 60/598,554
; PRIOR FILING DATE: 2004-08-03
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 64
; LENGTH: 11899
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-189-279-64

Query Match      54.2%; Score 18; DB 7; Length 118899;
Best Local Similarity 50.8%; Pred. No. 81;
Matches 21; Conservative 5; Mismatches 16; Indels 0; Gaps 0;

QY 1 GCNCAYGCNTGGCCNTCCNTAYAARGAYTAYGARGTNAAR 42
Db 11890 GGCAGGCCAGGCATTCACTTTATATGTTCTGCAGAGAAA 11849
```

```
RESULT 8
US-11-217-529-5102
; Sequence 5102, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: Patentin version 3.3
; SEQ ID NO 5102
; LENGTH: 2676
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-5102

Query Match      53.6%; Score 17.8; DB 7; Length 2676;
Best Local Similarity 50.0%; Pred. No. 29;
Matches 19; Conservative 6; Mismatches 13; Indels 0; Gaps 0;

QY 5 AYCNTGGCCNTCCNTAYAARGAYTAYGARGTNAAR 42
Db 379 ATGGCTGGCCATGTCTCATCAAGAACAACACGCTCAAG 416
```

```
RESULT 9
US-10-953-349-38683
; Sequence 38683, Application US/109533349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE OF INVENTION: ENCODED THERBY
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 38683
; LENGTH: 818
; TYPE: DNA
; ORGANISM: Zea mays subsp. mays
US-10-953-349-38683

Query Match          53.0%; Score 17.6; DB 6; Length 818;
Best Local Similarity 47.6%; Pred. No. 24;
Matches 20; Conservative 6; Mismatches 16; Indels 0; Gaps 0;

QY 1 GCNAYGCGTGGCCNTCNCNTAYAARGAYTAGGTNAAR 42
Db 439 GCCTTCGCGCGCGCGCGAGCTCAAGGAGTAGCAACAGAAG 480

RESULT 10
US-11-217-529-2422
; Sequence 2422, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHIISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 2422
; LENGTH: 939
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-2422

Query Match          53.0%; Score 17.6; DB 7; Length 939;
Best Local Similarity 51.3%; Pred. No. 25;
Matches 20; Conservative 5; Mismatches 14; Indels 0; Gaps 0;

QY 4 CAYGCGTGGCCNTCNCNTAYAARGAYTAGGTNAAR 42
Db 32 CAAAAATTGGCTAAGACCCTACAAGATTGGAAGTAAAA 70

RESULT 11
US-11-318-813-42/c
; Sequence 42, Application US/11318813
; Publication No. US20060105381A1
; GENERAL INFORMATION:
; APPLICANT: Ellipseis Biotherapeutics Corporation
; APPLICANT: Peltekova, Vanya D
; APPLICANT: Siminovich, Katherine A
; APPLICANT: St George-Hyslop, Peter H
```

```
; APPLICANT: Rubin, Laurence A
; APPLICANT: Peltekova, Vanya D
; APPLICANT: Wintle, Richard F
; TITLE OF INVENTION: POLYMORPHISMS OF THE OCTN1 AND OCTN2 CATION TRANSPORTERS ASSOCIATED WITH INFLAMMATORY BOWEL DISORDERS
; FILE OF INVENTION: INFLAMMATORY BOWEL DISORDERS
; FILE REFERENCE: ELLP-020
; CURRENT APPLICATION NUMBER: US/11/318,813
; CURRENT FILING DATE: 2005-12-27
; PRIOR APPLICATION NUMBER: US/10/327,188
; PRIOR FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: 60/362,700
; PRIOR FILING DATE: 2002-03-08
; PRIOR APPLICATION NUMBER: 60/343,338
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/427,529
; PRIOR FILING DATE: 2002-11-19
; PRIOR APPLICATION NUMBER: 60/362,717
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 42
; LENGTH: 54550
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (351)..(351)
; OTHER INFORMATION: n can be a or t or g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (436)..(436)
; OTHER INFORMATION: n can be a or t or g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (935)..(935)
; OTHER INFORMATION: n can be a or t or g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1725)..(1725)
; OTHER INFORMATION: n can be a or t or g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1874)..(1874)
; OTHER INFORMATION: n can be a or t or g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2031)..(2031)
; OTHER INFORMATION: n can be a or t or g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2152)..(2152)
; OTHER INFORMATION: n can be a or t or g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2297)..(2297)
; OTHER INFORMATION: n can be a or t or g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2747)..(2747)
; OTHER INFORMATION: n can be a or t or g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3477)..(3477)
; OTHER INFORMATION: n can be a or t or g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3535)..(3535)
; OTHER INFORMATION: n can be a or t or g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4035)..(4035)
; OTHER INFORMATION: n can be a or t or g or c
; FEATURE:
```



```
; ORGANISM: Mus musculus
US-10-488-619-3026

Query Match          52.4%; Score 17.4; DB 6; Length 955;
Best Local Similarity 55.3%; Pred. No. 31;
Matches 21; Conservative 3; Mismatches 14; Indels 0; Gaps 0;

QY 5 AYCNTGGCCNTCCNTAYARGAYTAYGARGTNAAR 42
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3 AGGTTTGGCTCGCCTTATAATTAATTAGAGTAAAA 40

RESULT 13
US-10-471-571A-5437
; Sequence 5437, Application US/10471571A
; Publication No. US20060115490A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: F026927WO
; CURRENT APPLICATION NUMBER: US/10/471,571A
; CURRENT FILING DATE: 2003-09-12
; PRIOR APPLICATION NUMBER: GB-0107661.1
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 5642
; SOFTWARE: Seqwin99, version 1.03
; SEQ ID NO 5437
; LENGTH: 1269
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-10-471-571A-5437

Query Match          52.4%; Score 17.4; DB 6; Length 1269;
Best Local Similarity 47.4%; Pred. No. 34;
Matches 18; Conservative 7; Mismatches 13; Indels 0; Gaps 0;

QY 5 AYCNTGGCCNTCCNTAYARGAYTAYGARGTNAAR 42
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 701 ATGAATTTACTTACCGTTATAAGATTATGAATTAGAA 738

RESULT 14
US-11-217-529-76209
; Sequence 76209, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 76209
; LENGTH: 1674
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-76209

Query Match          52.4%; Score 17.4; DB 7; Length 1674;
Best Local Similarity 50.0%; Pred. No. 38;
Matches 18; Conservative 6; Mismatches 12; Indels 0; Gaps 0;

QY 7 GCNTGGCCNTCCNTAYARGAYTAYGARGTNAAR 42
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 196 GCGTCGCAACCACTACCAGGACAACGGGGTCAAG 231

RESULT 15
US-11-217-529-3913
; Sequence 3913, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 3913
; LENGTH: 1731
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-3913

Query Match          52.4%; Score 17.4; DB 7; Length 1731;
Best Local Similarity 50.0%; Pred. No. 38;
Matches 18; Conservative 6; Mismatches 12; Indels 0; Gaps 0;

QY 7 GCNTGGCCNTCCNTAYARGAYTAYGARGTNAAR 42
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 256 GCATCGCACACCCGTACCAGGACAACGGGGTCAAG 291

RESULT 16
US-10-505-928-269
; Sequence 269, Application US/10505928
; Publication No. US20060088532A1
; GENERAL INFORMATION:
; APPLICANT: Ludwig Institute for Cancer Research et al.
; TITLE OF INVENTION: LYMPHATIC ENDOTHELIAL GENES
; FILE REFERENCE: 28967/39178
; CURRENT APPLICATION NUMBER: US/10/505,928
; CURRENT FILING DATE: 2004-08-27
; PRIOR APPLICATION NUMBER: US 60/363,019
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 866
; SOFTWARE: PatentIn 3.2
; SEQ ID NO 269
; LENGTH: 1897
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-505-928-269

Query Match          52.4%; Score 17.4; DB 6; Length 1897;
Best Local Similarity 64.3%; Pred. No. 39;
Matches 18; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 2 CMCAYCGTGGCCNTCCNTAYARGA 29
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 946 CCAGGCGCTGGCCCTCCCGATACAGAGA 973

RESULT 17
US-11-293-697-1526/C
; Sequence 1526, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: H1-A0106
```



;  
; CURRENT APPLICATION NUMBER: US/11/293,697  
; CURRENT FILING DATE: 2005-12-05  
; PRIOR FILING DATE: US/10/108,260  
; PRIOR FILING DATE: 2002-03-28  
; NUMBER OF SEQ ID NOS: 5458  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1526  
; LENGTH: 2876  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-293-697-1526

Query Match 52.4%; Score 17.4; DB 7; Length 2876;  
Best Local Similarity 58.1%; Pred. No. 45;  
Matches 18; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

Qy 5 AYCNGTGGCCNTCCNTAYAAAGAYTAYGA 35  
Db 920 ATGATTGGCCACCCCTTAAAGAACTGCA 890

RESULT 18  
US-10-471-571A-523  
; Sequence 523, Application US/10471571A  
; Publication No. US20060115490A1  
; GENERAL INFORMATION:  
; APPLICANT: CHIRON SpA  
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS PROTEINS AND NUCLEIC ACIDS  
; FILE REFERENCE: P026927W0  
; CURRENT APPLICATION NUMBER: US/10/471,571A  
; PRIOR FILING DATE: 2003-09-12  
; PRIOR APPLICATION NUMBER: GB-0107661.1  
; PRIOR FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 5642  
; SOFTWARE: Seqwin99, version 1.03  
; SEQ ID NO 523  
; LENGTH: 972  
; TYPE: DNA  
; ORGANISM: Staphylococcus aureus  
US-10-471-571A-523

Query Match 51.8%; Score 17.2; DB 6; Length 972;  
Best Local Similarity 48.7%; Pred. No. 39;  
Matches 19; Conservative 6; Mismatches 14; Indels 0; Gaps 0;

Qy 4 CAYCNGTGGCCNTCCNTAYAAAGAYTAYGARGTNAAR 42  
Db 877 CATTTCATTACATCATCATTTAAGAACTTAGAAGATAAA 915

RESULT 19  
US-10-953-349-23606  
; Sequence 23606, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 23606  
; LENGTH: 1121  
; TYPE: DNA  
; ORGANISM: Glycine max  
US-10-953-349-23606

Query Match 51.8%; Score 17.2; DB 6; Length 1121;  
Best Local Similarity 55.9%; Pred. No. 41;  
Matches 19; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

Qy 4 CAYCNGTGGCCNTCCNTAYAAAGAYTAYGARG 37  
Db 642 CATGGATGGACTTTGGCGGTACAATGATGATG 675

RESULT 20  
US-11-217-529-78099  
; Sequence 78099, Application US/11217529  
; Publication No. US20060099612A1  
; GENERAL INFORMATION:  
; APPLICANT: SUNTORY LIMITED  
; APPLICANT: NAKAO, YOSHIHIRO  
; APPLICANT: NAKAMURA, NORIHIISA  
; APPLICANT: KODAMA, YUKIKO  
; APPLICANT: FUJIMURA, TOMOKO  
; APPLICANT: ASHIKARI, TOSHIHIKO  
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS  
; FILE REFERENCE: S-38-285  
; CURRENT APPLICATION NUMBER: US/11/217,529  
; CURRENT FILING DATE: 2005-09-02  
; PRIOR APPLICATION NUMBER: US 10/932,182  
; PRIOR FILING DATE: 2004-09-02  
; NUMBER OF SEQ ID NOS: 197023  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 78099  
; LENGTH: 1377  
; TYPE: DNA  
; ORGANISM: Saccharomyces pastorianus  
US-11-217-529-78099

Query Match 51.8%; Score 17.2; DB 7; Length 1377;  
Best Local Similarity 55.9%; Pred. No. 44;  
Matches 19; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

Qy 4 CAYCNGTGGCCNTCCNTAYAAAGAYTAYGARG 37  
Db 1180 CATGACAGGCCAGCTCCATATAAGGAAGATAATG 1213

RESULT 21  
US-10-953-349-29678/c  
; Sequence 29678, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 29678  
; LENGTH: 1349  
; TYPE: DNA  
; ORGANISM: Triticum aestivum  
US-10-953-349-29678

Query Match 51.2%; Score 17; DB 6; Length 1349;  
Best Local Similarity 56.7%; Pred. No. 54;  
Matches 17; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

Qy 2 CNCAYCNGTGGCCNTCCNTAYAAAGAYT 31  
Db 638 CTCCTGCGAGGTCTCTCTCTTATCAGGATT 609

RESULT 22  
US-11-217-529-754/c  
; Sequence 754, Application US/11217529  
; Publication No. US20060099612A1  
; GENERAL INFORMATION:  
; APPLICANT: SUNTORY LIMITED

```

; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAWA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 754
; LENGTH: 1731
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-754

Query Match      51.2%; Score 17; DB 7; Length 1731;
Best Local Similarity 54.8%; Pred. No. 59;
Matches 17; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY      7 GNTGGCCNTCCNTAYAARGAYTAYGARG 37
Db      358 GCTTGGCAATCATGTTATAGCACCGAAG 328

RESULT 23
US-11-217-529-1835/c
; Sequence 1835, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAWA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1835
; LENGTH: 1800
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-1835

Query Match      51.2%; Score 17; DB 7; Length 1800;
Best Local Similarity 60.7%; Pred. No. 60;
Matches 17; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY      11 GGCCNTCCNTAYAARGAYTAYGARGT 38
Db      1087 GGCCITCTTCCTTAAGGATGAAGAAGT 1060

RESULT 24
US-11-293-697-51/c
; Sequence 51, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
```

```

; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 51
; LENGTH: 2269
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-293-697-51

Query Match      51.2%; Score 17; DB 7; Length 2269;
Best Local Similarity 54.8%; Pred. No. 65;
Matches 17; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY      5 AYGCNTGGCCNTCCNTAYAARGAYTAYGA 35
Db      402 ATGCCTGGACTTCCAAAGTATAAATATCTGA 372

RESULT 25
US-11-293-697-2015/c
; Sequence 2015, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2015
; LENGTH: 2856
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-293-697-2015

Query Match      51.2%; Score 17; DB 7; Length 2856;
Best Local Similarity 54.8%; Pred. No. 70;
Matches 17; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY      7 GNTGGCCNTCCNTAYAARGAYTAYGARG 37
Db      2678 GCGTGTCTTCACACACACAGACAGTGAAG 2648

RESULT 26
US-10-953-349-31575
; Sequence 31575, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 31575
; LENGTH: 3007
; TYPE: DNA
; ORGANISM: Triticum aestivum
US-10-953-349-31575

Query Match      51.2%; Score 17; DB 6; Length 3007;
Best Local Similarity 60.7%; Pred. No. 71;
Matches 17; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY      11 GGCCNTCCNTAYAARGAYTAYGARGT 38
Db      757 GGCCATCAGATATAGGATTATCATGT 784
```

```
RESULT 27
US-11-293-697-1736/c
; Sequence 1736, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: HI-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; PRIOR FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1736
; LENGTH: 3008
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-293-697-1736

Query Match          51.2%; Score 17; DB 7; Length 3008;
Best Local Similarity 54.8%; Pred. No. 71;
Matches 17; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Qy      7 GCNTGGCCNTCNCNTAYAAAGAYTAYGARG 37
Db      2832 GCGTGGTCTTCACACACACAGAGTGAAG 2802

RESULT 28
US-10-502-993-1/c
; Sequence 1, Application US/10502993
; Publication No. US20060089320A1
; GENERAL INFORMATION:
; APPLICANT: CANTLEY, Lewis C.
; APPLICANT: LAMIA, Katja A.
; APPLICANT: RAMEH, Lucia
; APPLICANT: KAHN, Barbara
; APPLICANT: PERONI, Odile
; TITLE OF INVENTION: MODULATION OF TYPE IIa PHOSPHONOSITIDE PHOSPHATE KINASE
; FILE REFERENCE: B0662.70052U01
; CURRENT APPLICATION NUMBER: US/10/502,993
; CURRENT FILING DATE: 2004-07-29
; PRIOR APPLICATION NUMBER: US 60/353,758
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 3743
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (481)..(1728)
US-10-502-993-1

Query Match          51.2%; Score 17; DB 6; Length 3743;
Best Local Similarity 51.3%; Pred. No. 77;
Matches 20; Conservative 4; Mismatches 15; Indels 0; Gaps 0;

Qy      2 CNCAYGCGTGGCCNTCNCNTAYAAAGAYTAYGARGTNA 40
Db      2480 CGCTTTCCTCCCTCCACCATCCCATGAGGATGAACCTTA 2442

RESULT 29
US-11-217-529-3130
; Sequence 3130, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
```

```
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 3130
; LENGTH: 4656
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-3130

Query Match          51.2%; Score 17; DB 7; Length 4656;
Best Local Similarity 51.5%; Pred. No. 83;
Matches 17; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

Qy      4 CAYGCGTGGCCNTCNCNTAYAAAGAYTAYGAR 36
Db      538 CATGAATTGCTTTGTCTTATAAAGACTTTGAG 570

RESULT 30
US-10-473-173-112
; Sequence 112, Application US/10473173
; Publication No. US2006008823A1
; GENERAL INFORMATION:
; APPLICANT: VAN ANDEL INSTITUTE
; TITLE OF INVENTION: Microarray Gene Expression Profiling in Clear Cell Renal Cell
; FILE REFERENCE: 38345-170094
; CURRENT APPLICATION NUMBER: US/10/473,173
; CURRENT FILING DATE: 2003-09-29
; PRIOR APPLICATION NUMBER: US 60/279,411
; PRIOR FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 498
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 112
; LENGTH: 5003
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-473-173-112

Query Match          51.2%; Score 17; DB 6; Length 5003;
Best Local Similarity 52.6%; Pred. No. 85;
Matches 20; Conservative 4; Mismatches 14; Indels 0; Gaps 0;

Qy      1 GCNCAYGCGTGGCCNTCNCNTAYAAAGAYTAYGARGT 38
Db      2214 GAACCTGCGCTGCTTCTCCCAAAGGACAAAGAGGT 2251

RESULT 31
US-10-953-349-679
; Sequence 679, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 679
; LENGTH: 802
```



RESULT 36  
US-11-318-813-42  
; Sequence 42, Application US/11318813  
; Publication No. US20060105381A1  
; GENERAL INFORMATION:  
; APPLICANT: Ellipsis Biotherapeutics Corporation  
; APPLICANT: Peltekova, Vanya D  
; APPLICANT: Siminovich, Katherine A  
; APPLICANT: St George-Hyslop, Peter H  
; APPLICANT: Rubin, Laurence A  
; APPLICANT: Peltekova, Vanya D  
; APPLICANT: Wintle, Richard F  
; TITLE OF INVENTION: POLYMORPHISMS OF THE OCTN1 AND OCTN2 CATION TRANSPORTERS ASSOCIATED WITH INFLAMMATORY BOWEL DISORDERS  
; FILE REFERENCE: ELLP-020  
; CURRENT APPLICATION NUMBER: US/11/318,813  
; CURRENT FILING DATE: 2005-12-27  
; PRIOR APPLICATION NUMBER: US/10/327,188  
; PRIOR FILING DATE: 2002-12-20  
; PRIOR APPLICATION NUMBER: 60/362,700  
; PRIOR FILING DATE: 2002-03-08  
; PRIOR APPLICATION NUMBER: 60/343,338  
; PRIOR FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: 60/427,529  
; PRIOR FILING DATE: 2002-11-19  
; PRIOR APPLICATION NUMBER: 60/362,717  
; PRIOR FILING DATE: 2002-03-08  
; NUMBER OF SEQ ID NOS: 42  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 42  
; LENGTH: 54550  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (351)..(351)  
; OTHER INFORMATION: n can be a or t or g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (436)..(436)  
; OTHER INFORMATION: n can be a or t or g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (935)..(935)  
; OTHER INFORMATION: n can be a or t or g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1725)..(1725)  
; OTHER INFORMATION: n can be a or t or g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1874)..(1874)  
; OTHER INFORMATION: n can be a or t or g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (2031)..(2031)  
; OTHER INFORMATION: n can be a or t or g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (2152)..(2152)  
; OTHER INFORMATION: n can be a or t or g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (2297)..(2297)  
; OTHER INFORMATION: n can be a or t or g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (2747)..(2747)  
; OTHER INFORMATION: n can be a or t or g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (3477)..(3477)  
; OTHER INFORMATION: n can be a or t or g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (3535)..(3535)  
; OTHER INFORMATION: n can be a or t or g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (4035)..(4035)  
; OTHER INFORMATION: n can be a or t or g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (4084)..(4084)  
; OTHER INFORMATION: n can be a or t or g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (4087)..(4087)  
; OTHER INFORMATION: n can be a or t or g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (4321)..(4321)  
; OTHER INFORMATION: n can be a or t or g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (4374)..(4374)  
; OTHER INFORMATION: n can be a or t or g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (4394)..(4394)  
; OTHER INFORMATION: n can be a or t or g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (4429)..(4429)  
; OTHER INFORMATION: n can be a or t or g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (4480)..(4480)  
; OTHER INFORMATION: n can be a or t or g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (4482)..(4482)  
; OTHER INFORMATION: n can be a or t or g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (4704)..(4704)  
; OTHER INFORMATION: n can be a or t or g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (4799)..(4799)  
; OTHER INFORMATION: n can be a or t or g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (4800)..(4800)  
; OTHER INFORMATION: n can be a or t or g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (4801)..(4801)  
; OTHER INFORMATION: n can be a or t or g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (4802)..(4802)  
; OTHER INFORMATION: n can be a or t or g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (4803)..(4803)  
; OTHER INFORMATION: n can be a or t or g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (4805)..(4805)  
; OTHER INFORMATION: n can be a or t or g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (4806)..(4806)  
; OTHER INFORMATION: n can be a or t or g or c



```
Query Match      50.0%; Score 16.6; DB 6; Length 644;
Best Local Similarity 52.8%; Pred. No. 65;
Matches 19; Conservative 4; Mismatches 13; Indels 0; Gaps 0;

Qy      6 YGNTGGCCNTCCNTAARGAYTAYGARGTNA 41
Db      330 TGCTATGCCGCCCTATAGAAGACCGATGACA 295

RESULT 40
US-10-488-619-2143
; Sequence 2143, Application US/10488619
; Publication No. US2006009578A1
; GENERAL INFORMATION:
; APPLICANT: Greenlee, Winner and Sullivan, P.C.
; TITLE OF INVENTION: Human Mitochondrial DNA Polymorphisms, Haplogroups, Associations
; FILE OF INVENTION: Physiological Conditions, And Genotyping Arrays
; FILE REFERENCE: 98-01 WO
; CURRENT APPLICATION NUMBER: US/10/488,619
; CURRENT FILING DATE: 2004-03-01
; NUMBER OF SEQ ID NOS: 3040
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2143
; LENGTH: 851
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-488-619-2143

Query Match      50.0%; Score 16.6; DB 6; Length 851;
Best Local Similarity 50.0%; Pred. No. 71;
Matches 19; Conservative 5; Mismatches 14; Indels 0; Gaps 0;

Qy      4 CAYCNTGGCCNTCCNTAARGAYTAYGARGTNA 41
Db      657 CATCCAGGCCCTTCTCTGATAAGTCTTGGAGGATA 694

RESULT 41
US-10-953-349-31441
; Sequence 31441, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 31441
; LENGTH: 1266
; TYPE: DNA
; ORGANISM: Triticum aestivum
US-10-953-349-31441

Query Match      50.0%; Score 16.6; DB 6; Length 1266;
Best Local Similarity 50.0%; Pred. No. 82;
Matches 19; Conservative 5; Mismatches 14; Indels 0; Gaps 0;

Qy      1 GCNCAVCGNTGCCNTCCNTAARGAYTAYGARGCT 38
Db      702 GACACGGTGGCCGCCGTGTGGACGCGATACCGGT 739

RESULT 42
US-10-953-349-25421/c
; Sequence 25421, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
```

```
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 25421
; LENGTH: 1312
; TYPE: DNA
; ORGANISM: Triticum aestivum
US-10-953-349-25421

Query Match      50.0%; Score 16.6; DB 6; Length 1312;
Best Local Similarity 55.9%; Pred. No. 83;
Matches 19; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

Qy      7 GNTGGCCNTCCNTAARGAYTAYGARGTNA 40
Db      241 GCGATGCCGCCCGGTAGAAAGGACCGAGGTGA 208

RESULT 43
US-10-953-349-8766/c
; Sequence 8766, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 8766
; LENGTH: 1746
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-10-953-349-8766

Query Match      50.0%; Score 16.6; DB 6; Length 1746;
Best Local Similarity 46.3%; Pred. No. 91;
Matches 19; Conservative 6; Mismatches 16; Indels 0; Gaps 0;

Qy      2 CNCAYCGNTGCCNTCCNTAARGAYTAYGARGTNA 42
Db      470 CAGACACTTGTCTCTACCAACACGCGATCTGGAGTGA 430

RESULT 44
US-11-217-529-76864
; Sequence 76864, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 76864
; LENGTH: 2106
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-76864
```

Query Match 50.0%; Score 16.6; DB 7; Length 2106;  
Best Local Similarity 46.3%; Pred. No. 97;  
Matches 19; Conservative 6; Mismatches 16; Indels 0; Gaps 0;

QY 1 GCNCGCNCNTGCCNTCNCNTAYAAARGAYTAYGARGTNA 41  
Db 229 GGCACCTCTTTGCCCTCTGAGGATGAAGACATCGAAGTTAA 269

RESULT 45  
US-11-217-529-78513  
; Sequence 78513, Application US/11217529  
; Publication No. US20060099612A1  
; GENERAL INFORMATION:  
; APPLICANT: SUNTORY LIMITED  
; APPLICANT: NAKAO, YOSHIHIRO  
; APPLICANT: NAKAMURA, NORIHISA  
; APPLICANT: KODAMA, YUKIO  
; APPLICANT: FUJIMURA, TOMOKO  
; APPLICANT: ASHIKARI, TOSHIHIKO  
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS  
; FILE REFERENCE: S-38-285  
; CURRENT APPLICATION NUMBER: US/11/217,529  
; CURRENT FILING DATE: 2005-09-02  
; PRIOR APPLICATION NUMBER: US 10/932,182  
; PRIOR FILING DATE: 2004-09-02  
; NUMBER OF SEQ ID NOS: 197023  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 78513  
; LENGTH: 3306  
; TYPE: DNA  
; ORGANISM: Saccharomyces pastorianus  
US-11-217-529-78513

Query Match 50.0%; Score 16.6; DB 7; Length 3306;  
Best Local Similarity 52.8%; Pred. No. 1.1e+02;  
Matches 19; Conservative 4; Mismatches 13; Indels 0; Gaps 0;

QY 7 GCNTGGCNCNTCNCNTAYAAARGAYTAYGARGTNAAR 42  
Db 736 GGTTCGAAATCGCAATATAATGAATATGAGGAGAA 771

RESULT 46  
US-10-953-349-13812  
; Sequence 13812, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; FILE REFERENCE: ENCODED THERBY  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 13812  
; LENGTH: 530  
; TYPE: DNA  
; ORGANISM: Glycine max  
US-10-953-349-13812

Query Match 49.4%; Score 16.4; DB 6; Length 530;  
Best Local Similarity 50.0%; Pred. No. 75;  
Matches 20; Conservative 4; Mismatches 16; Indels 0; Gaps 0;

QY 2 CNCAYGNCNTGCCNTCNCNTAYAAARGAYTAYGARGTNA 41  
Db 198 CGAATGCGTCATGTGTTCTATATATACAGTATGATGTCAA 237

RESULT 47  
US-10-488-619-1977/c

; Sequence 1977, Application US/10488619  
; Publication No. US20060099578A1  
; GENERAL INFORMATION:  
; APPLICANT: Greenlee, Winner and Sullivan, P.C.  
; TITLE OF INVENTION: Human Mitochondrial DNA Polymorphisms, Haplogroups, Associations  
; FILE REFERENCE: 98-01 WO  
; CURRENT APPLICATION NUMBER: US/10/488,619  
; CURRENT FILING DATE: 2004-03-01  
; NUMBER OF SEQ ID NOS: 3040  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1977  
; LENGTH: 698  
; TYPE: DNA  
; ORGANISM: Mus musculus  
US-10-488-619-1977

Query Match 49.4%; Score 16.4; DB 6; Length 698;  
Best Local Similarity 57.1%; Pred. No. 83;  
Matches 20; Conservative 2; Mismatches 13; Indels 0; Gaps 0;

QY 1 GCNCGCNCNTGCCNTCNCNTAYAAARGAYTAYGA 35  
Db 521 GCCAAGCCTGCTCCTGTCTGTCAAATGACTATAA 487

RESULT 48  
US-10-488-619-1795  
; Sequence 1795, Application US/10488619  
; Publication No. US20060099578A1  
; GENERAL INFORMATION:  
; APPLICANT: Greenlee, Winner and Sullivan, P.C.  
; TITLE OF INVENTION: Human Mitochondrial DNA Polymorphisms, Haplogroups, Associations  
; FILE REFERENCE: 98-01 WO  
; CURRENT APPLICATION NUMBER: US/10/488,619  
; CURRENT FILING DATE: 2004-03-01  
; NUMBER OF SEQ ID NOS: 3040  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1795  
; LENGTH: 909  
; TYPE: DNA  
; ORGANISM: Mus musculus  
US-10-488-619-1795

Query Match 49.4%; Score 16.4; DB 6; Length 909;  
Best Local Similarity 56.7%; Pred. No. 91;  
Matches 17; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 4 CAYGNCNTGCCNTCNCNTAYAAARGAYTAY 33  
Db 634 CAGGATTGCCATCTCCAGCCACAGACTAC 663

RESULT 49  
US-10-953-349-39951/c  
; Sequence 39951, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; FILE REFERENCE: ENCODED THERBY  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 39951  
; LENGTH: 1087  
; TYPE: DNA  
; ORGANISM: Zea mays subsp. mays  
US-10-953-349-39951



```
Query Match          49.4%; Score 16.4; DB 6; Length 1087;
Best Local Similarity 47.6%; Pred. No. 96;
Matches 20; Conservative 5; Mismatches 17; Indels 0; Gaps 0;

Qy      1 GCNAYGNTGCCNTCCNTAYAAARGAYTAYGARGTNAAR 42
Db      320 GATCAAGGGTGATCTTCTCTTACGGAGAATATCAACTGCAG 279

RESULT 50
US-10-953-349-24567/c
; Sequence 24567, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 24567
; LENGTH: 1228
; TYPE: DNA
; ORGANISM: Glycine max
US-10-953-349-24567

Query Match          49.4%; Score 16.4; DB 6; Length 1228;
Best Local Similarity 50.0%; Pred. No. 1e+02;
Matches 20; Conservative 4; Mismatches 16; Indels 0; Gaps 0;

Qy      2 CNCAYGNTGCCNTCCNTAYAAARGAYTAYGARGTNA 41
Db      1038 CTATGCTTGTCCATCACTTAACCAATTATTATTGAA 999

RESULT 51
US-10-953-349-35818/c
; Sequence 35818, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 35818
; LENGTH: 1406
; TYPE: DNA
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (534)..(534)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (659)..(659)
; OTHER INFORMATION: n is a, c, g, or t
US-10-953-349-35818

Query Match          49.4%; Score 16.4; DB 6; Length 1406;
Best Local Similarity 54.8%; Pred. No. 1.1e+02;
Matches 17; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

Qy      2 CNCAYGNTGCCNTCCNTAYAAARGAYTA 32
Db      941 CGCATGCATGCATCATCTGCTATAAATACGA 911
```

```
RESULT 52
US-10-953-349-10383
; Sequence 10383, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 10383
; LENGTH: 1462
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-10-953-349-10383

Query Match          49.4%; Score 16.4; DB 6; Length 1462;
Best Local Similarity 48.6%; Pred. No. 1.1e+02;
Matches 17; Conservative 6; Mismatches 12; Indels 0; Gaps 0;

Qy      6 YGNTGGCCNTCCNTAYAAARGAYTAYGARGTNA 40
Db      1216 CGCTTGGTCTCCTCCATGTTAGAACTTTGAAGAGA 1250

RESULT 53
US-11-217-529-79698/c
; Sequence 79698, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 79698
; LENGTH: 1566
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-79698

Query Match          49.4%; Score 16.4; DB 7; Length 1566;
Best Local Similarity 60.7%; Pred. No. 1.1e+02;
Matches 17; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

Qy      8 CNTGGCCNTCCNTAYAAARGAYTAYGA 35
Db      1404 CATTTCCTTCACCACAAAAGATTACGA 1377

RESULT 54
US-11-293-697-2093
; Sequence 2093, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
```

```
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2093
; LENGTH: 1854
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-293-697-2093

Query Match          49.4%; Score 16.4; DB 7; Length 1854;
Best Local Similarity 47.6%; Pred. No. 1.2e+02;
Matches 20; Conservative 5; Mismatches 17; Indels 0; Gaps 0;

QY      1 GCNCAYGCGTGGCCNTCNCNTAYAAARGAYTAYGARGTNAAR 42
Db      1460 GCTCAGGCGCTGGACCCACCGTGACAGGCTGCTGAGGGTAAG 1501

RESULT 55
US-10-505-928-771
; Sequence 771, Application US/10505928
; Publication No. US20060088532A1
; GENERAL INFORMATION:
; APPLICANT: Ludwig Institute for Cancer Research et al.
; TITLE OF INVENTION: LYMPHATIC ENDOTHELIAL GENES
; FILE REFERENCE: 28967/33178
; CURRENT APPLICATION NUMBER: US/10/505,928
; CURRENT FILING DATE: 2004-08-27
; PRIOR APPLICATION NUMBER: US 60/363,019
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 866
; SOFTWARE: PatentIn 3.2
; SEQ ID NO 771
; LENGTH: 2180
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-505-928-771

Query Match          49.4%; Score 16.4; DB 6; Length 2180;
Best Local Similarity 47.6%; Pred. No. 1.2e+02;
Matches 20; Conservative 5; Mismatches 17; Indels 0; Gaps 0;

QY      1 GCNCAYGCGTGGCCNTCNCNTAYAAARGAYTAYGARGTNAAR 42
Db      1724 GCTCAGGCGCTGGACCCACCGTGACAGGCTGCTGAGGGTAAG 1765

RESULT 56
US-11-217-529-76276/c
; Sequence 76276, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 76276
; LENGTH: 2235
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-76276

Query Match          49.4%; Score 16.4; DB 7; Length 2235;
Best Local Similarity 47.6%; Pred. No. 1.3e+02;
Matches 20; Conservative 5; Mismatches 17; Indels 0; Gaps 0;

QY      1 GCNCAYGCGTGGCCNTCNCNTAYAAARGAYTAYGARGTNAAR 42
Db      1079 GCTCAGGCGCTGGACCCACCGTGACAGGCTGCTGAGGGTAAG 1120

RESULT 59
US-10-505-928-554/c
; Sequence 554, Application US/10505928
; Publication No. US20060088532A1
; GENERAL INFORMATION:
; APPLICANT: Ludwig Institute for Cancer Research et al.
; TITLE OF INVENTION: LYMPHATIC ENDOTHELIAL GENES

Best Local Similarity 56.7%; Pred. No. 1.2e+02;
Matches 17; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY      11 GGCCNTCNCNTAYAAARGAYTAYGARGTNA 40
Db      320 GGCCATCTCTTGCAGAAAACATGAAGTGA 291

RESULT 57
US-11-293-697-834
; Sequence 834, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 834
; LENGTH: 2426
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-293-697-834

Query Match          49.4%; Score 16.4; DB 7; Length 2426;
Best Local Similarity 50.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 4; Mismatches 16; Indels 0; Gaps 0;

QY      1 GCNCAYGCGTGGCCNTCNCNTAYAAARGAYTAYGARGTNA 40
Db      1125 GCATGTGCTCTCTCTTAAGTGAAGACTATGACGACA 1164

RESULT 58
US-11-293-697-289
; Sequence 289, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 289
; LENGTH: 2700
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-293-697-289

Query Match          49.4%; Score 16.4; DB 7; Length 2700;
Best Local Similarity 47.6%; Pred. No. 1.3e+02;
Matches 20; Conservative 5; Mismatches 17; Indels 0; Gaps 0;

QY      1 GCNCAYGCGTGGCCNTCNCNTAYAAARGAYTAYGARGTNAAR 42
Db      1079 GCTCAGGCGCTGGACCCACCGTGACAGGCTGCTGAGGGTAAG 1120

RESULT 59
US-10-505-928-554/c
; Sequence 554, Application US/10505928
; Publication No. US20060088532A1
; GENERAL INFORMATION:
; APPLICANT: Ludwig Institute for Cancer Research et al.
; TITLE OF INVENTION: LYMPHATIC ENDOTHELIAL GENES
```





Qy	1	GNCAYGCNTGGCCNTCNCNTAYAARGAYTA	32
		:	
Db	24	GCCACACCTGGCCTTCTCCATGCTCGGAATA	55

```

RESULT 68
US-11-217-529-2174
; Sequence 2174, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHIISA
; APPLICANT: KODAMA, YUKIO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: 8-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 2174
; LENGTH: 1632
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-2174

```

```

RESULT 69
US-10-953-349-2142
; Sequence 2142, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; TITLE OF INVENTION: ENCODED THERBY
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: Patentin version 3.3
; SEQ ID NO 2142
; LENGTH: 1745
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-10-953-349-2142

Query Match 48.8%; Score 16.2; DB 6; Length 1745;
Best Local Similarity 57.7%; Pred. No. 1.4e+02;
Matches 15; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 16 TCNCCNTAYAAAGGAYTAYGAGGTNAA 41
|||:|:|:|:|:|:|:|:|:|
DB 599 TCAATCTACGAAGGACTACGAAGACAA 624

RESULT 70
US-10-953-349-9257/c
; Sequence 9257, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.

```

```

; ; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; ;
; ; TITLE OF INVENTION: ENCODED THERBY
; ;
; ; FILE REFERENCE: 2750-1579PUS2
; ;
; ; CURRENT APPLICATION NUMBER: US/10/953,349
; ;
; ; CURRENT FILING DATE: 2004-09-30
; ;
; ; NUMBER OF SEQ ID NOS: 40252
; ;
; ; SOFTWARE: PatentIn version 3.3
; ;
; ; SEQ ID NO 9257
; ;
; ; LENGTH: 1828
; ;
; ; TYPE: DNA
; ;
; ; ORGANISM: Arabidopsis thaliana
; ;
; ; US-10-953-349-9257

```

```

RESULT 71
US-10-953-349-2382/c
; Sequence 2382, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nickolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; TITLE OF INVENTION: ENCODED THERAPY
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 2382
; LENGTH: 1831
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-10-953-349-2382

```

RESULT 72  
US-11-217-4479  
; Sequence 4479, Application US/11217529  
; Publication No. US20060099612A1  
; GENERAL INFORMATION:  
; APPLICANT: SUNTORY LIMITED  
; APPLICANT: NAKAO, YOSHIHIRO  
; APPLICANT: NAKAMURA, NORIHIISA  
; APPLICANT: KODAMA, YUKIKO  
; APPLICANT: FUJIMURA, TOMOKO  
; APPLICANT: ASHIKARI, TOSHIHIKO  
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS  
; FILE REFERENCE: S-38-285  
; CURRENT APPLICATION NUMBER: US/11/217,529  
; CURRENT FILING DATE: 2005-09-02  
; PRIOR APPLICATION NUMBER: US 10/932,182  
; PRIOR FILING DATE: 2004-09-02  
; NUMBER OF SEQ ID NOS: 197023  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 4479  
; LENGTH: 2163  
; TYPE: DNA  
; ORGANISM: Saccharomyces pastorianus





Best Local Similarity 57.1%; Pred. No. 1.2e+02; Indels 0; Gaps 0;  
Matches 16; Conservative 4; Mismatches 8;

QY 13 CCNTGCCNTCAAGGATGACTCA 84  
DB 57 CCATGTCCTACAGGATGACTCA 84

RESULT 81

US-11-217-529-79889/c  
; Sequence 79889, Application US/11217529  
; Publication No. US20060099612A1  
; GENERAL INFORMATION:  
; APPLICANT: SUNTORY LIMITED  
; APPLICANT: NAKAO, YOSHIHIRO  
; APPLICANT: NAKAMURA, NORIHISA  
; APPLICANT: KODAMA, YUKIKO  
; APPLICANT: FUJIMURA, TOMOKO  
; APPLICANT: ASHIKARI, TOSHIHIKO  
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS  
; FILE REFERENCE: S-38-285  
; CURRENT APPLICATION NUMBER: US/11/217,529  
; CURRENT FILING DATE: 2005-09-02  
; PRIOR APPLICATION NUMBER: US 10/932,182  
; PRIOR FILING DATE: 2004-09-02  
; NUMBER OF SEQ ID NOS: 197023  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 79889  
; LENGTH: 735  
; TYPE: DNA  
; ORGANISM: Saccharomyces pastorianus  
US-11-217-529-79889

Query Match 48.2%; Score 16; DB 7; Length 735;  
Best Local Similarity 51.4%; Pred. No. 1.3e+02; Indels 0; Gaps 0;  
Matches 19; Conservative 4; Mismatches 14;

QY 6 YGNTGCCNTCCNTCAAGGATGAGTNAAR 42  
DB 151 TGCTTGGGATTCGCGTAAAGAGAAAGAAATTCAA 115

RESULT 82

US-10-471-571A-3165/c  
; Sequence 3165, Application US/10471571A  
; Publication No. US20060115490A1  
; GENERAL INFORMATION:  
; APPLICANT: CHIRON SPA  
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS PROTEINS AND NUCLEIC ACIDS  
; FILE REFERENCE: P026927WO  
; CURRENT APPLICATION NUMBER: US/10/471,571A  
; CURRENT FILING DATE: 2003-09-12  
; PRIOR APPLICATION NUMBER: GB-0107661.1  
; PRIOR FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 5642  
; SOFTWARE: SeqWin99, version 1.03  
; SEQ ID NO 3165  
; LENGTH: 762  
; TYPE: DNA  
; ORGANISM: Staphylococcus aureus  
US-10-471-571A-3165

Query Match 48.2%; Score 16; DB 6; Length 762;  
Best Local Similarity 51.4%; Pred. No. 1.3e+02; Indels 0; Gaps 0;  
Matches 19; Conservative 4; Mismatches 14;

QY 4 CAYGNTGCCNTCCNTCAAGGATGAGTNAAR 40  
DB 151 CAGGGTGGCTGTACTATATAAAGATATTTCGTTGA 115

RESULT 83

US-10-953-349-6898/c

; Sequence 6898, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; TITLE OF INVENTION: ENCODED THERBY  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 6898  
; LENGTH: 914  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-10-953-349-6898

Query Match 48.2%; Score 16; DB 6; Length 914;  
Best Local Similarity 45.2%; Pred. No. 1.4e+02; Indels 0; Gaps 0;  
Matches 19; Conservative 6; Mismatches 17;

QY 1 GCNCAYGNTGCCNTCCNTCAAGGATGAGTNAAR 42  
DB 728 GCTCATGCGTGGCGGACCCCTTTAAAAAGCCGAGTCAAG 687

RESULT 84

US-10-953-349-35524  
; Sequence 35524, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; TITLE OF INVENTION: ENCODED THERBY  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 35524  
; LENGTH: 1227  
; TYPE: DNA  
; ORGANISM: Zea mays subsp. mays  
US-10-953-349-35524

Query Match 48.2%; Score 16; DB 6; Length 1227;  
Best Local Similarity 45.2%; Pred. No. 1.5e+02; Indels 0; Gaps 0;  
Matches 19; Conservative 6; Mismatches 17;

QY 1 GCNCAYGNTGCCNTCCNTCAAGGATGAGTNAAR 42  
DB 251 GCTTACGCATGGCTGAGCCATAACACAGATGTTGATGACAAG 292

RESULT 85

US-10-953-349-29958  
; Sequence 29958, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; TITLE OF INVENTION: ENCODED THERBY  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 29958  
; LENGTH: 1358  
; TYPE: DNA  
; ORGANISM: Triticum aestivum  
; FEATURE:  
; NAME/KEY: misc\_feature



LOCATION: (962)..(962)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1010)..(1010)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1032)..(1032)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1080)..(1080)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1090)..(1090)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1152)..(1152)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1163)..(1163)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1228)..(1228)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1239)..(1239)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1331)..(1331)  
; OTHER INFORMATION: n is a, c, g, or t  
US-10-953-349-29958

Query Match 48.2%; Score 16; DB 6; Length 1358;  
Best Local Similarity 47.6%; Pred. No. 1.6e+02;  
Matches 20; Conservative 6; Mismatches 16; Indels 0; Gaps 0;

Qy 1 GCNCAYCGTCGCCNTCCNTAYAAAGAYTAYGARGTNAAR 42  
Db 945 GTTCATGAAGAGCGCGCCGCCCCCAAGGAAACAAAGGATAAG 986

RESULT 86  
US-10-953-349-21144/c  
; Sequence 21144, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: Patentin version 3.3  
; SEQ ID NO 21144  
; LENGTH: 1371  
; TYPE: DNA  
; ORGANISM: Glycine max  
US-10-953-349-21144

Query Match 48.2%; Score 16; DB 6; Length 1371;  
Best Local Similarity 55.2%; Pred. No. 1.6e+02;  
Matches 16; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

Qy 10 TGCCNTCCNTAYAAAGAYTAYGARGT 38

Db 242 TGGCCTTCATTCATTAAAGACTGTGAGGT 214

RESULT 87  
US-10-953-349-24287  
; Sequence 24287, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: Patentin version 3.3  
; SEQ ID NO 24287  
; LENGTH: 1499  
; TYPE: DNA  
; ORGANISM: Glycine max  
US-10-953-349-24287

Query Match 48.2%; Score 16; DB 6; Length 1499;  
Best Local Similarity 51.4%; Pred. No. 1.7e+02;  
Matches 19; Conservative 4; Mismatches 14; Indels 0; Gaps 0;

Qy 2 CNCAYCGTCGCCNTCCNTAYAAAGAYTAYGARGT 38  
Db 689 CTCACGCGCTTCAGTCACTCTACAACGATTCGGAATT 725

RESULT 88  
US-10-953-349-12909  
; Sequence 12909, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: Patentin version 3.3  
; SEQ ID NO 12909  
; LENGTH: 1584  
; TYPE: DNA  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (972)..(972)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1573)..(1573)  
; OTHER INFORMATION: n is a, c, g, or t  
US-10-953-349-12909

Query Match 48.2%; Score 16; DB 6; Length 1584;  
Best Local Similarity 47.5%; Pred. No. 1.7e+02;  
Matches 19; Conservative 5; Mismatches 16; Indels 0; Gaps 0;

Qy 2 CNCAYCGTCGCCNTCCNTAYAAAGAYTAYGARGTNA 41  
Db 37 CTCATGATGTCCTCTCTATCTTCTCCCAATGAATGAA 76

RESULT 89  
US-09-949-925-71/c  
; Sequence 71, Application US/09949925  
; Publication No. US20060099575A9  
; GENERAL INFORMATION:







GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2006, 15:26:46 ; Search time 1608.6 Seconds  
(without alignments)  
320.826 Million cell updates/sec

Title: US-10-600-816-33  
Perfect score: 33.2  
Sequence: 1 gncaycgtgcctcnc.....ayaargaytaygargtnaar 42

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 18892170 seqs, 6143817638 residues

Total number of hits satisfying chosen parameters: 37784340

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 1000 summaries

## Database : Published Applications NA Main:\*

1: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
3: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US09A\_PUBCOMB.seq:\*  
4: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*  
5: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*  
6: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*  
7: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*  
8: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*  
9: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10D\_PUBCOMB.seq:\*  
10: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10E\_PUBCOMB.seq:\*  
11: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10F\_PUBCOMB.seq:\*  
12: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10G\_PUBCOMB.seq:\*  
13: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US11A\_PUBCOMB.seq:\*  
14: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US11B\_PUBCOMB.seq:\*  
15: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US11C\_PUBCOMB.seq:\*  
16: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US11D\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	33.2	100.0	42	8	US-10-600-816-33
2	30	90.4	42	8	US-10-600-816-30
3	30	90.4	497	6	US-10-066-543-1937
C 4	30	90.4	552	6	US-10-066-543-1811
C 5	30	90.4	620	3	US-09-969-034-2222
C 6	30	90.4	634	3	US-09-969-034-3393
C 7	30	90.4	642	7	US-10-125-968-701
8	30	90.4	1071	9	US-10-712-615-134
9	30	90.4	1212	3	US-09-866-050A-249
10	30	90.4	1212	6	US-10-152-661-249
11	30	90.4	1228	7	US-10-313-542-223
12	30	90.4	1400	13	US-11-060-756-2418
13	30	90.4	1400	13	US-11-060-756-6690
14	30	90.4	1460	10	US-10-936-626-64
15	30	90.4	1460	10	US-10-938-061-64
16	30	90.4	1619	7	US-10-224-289-5
17	30	90.4	1619	9	US-10-935-190-43

18	30	90.4	1788	10	US-10-505-486-196	Sequence 196, App
19	30	90.4	2297	9	US-10-775-920-13	Sequence 13, Appli
20	30	90.4	2302	7	US-10-224-289-3	Sequence 3, Appli
21	30	90.4	2302	8	US-10-240-425-405	Sequence 405, App
22	30	90.4	2302	9	US-10-775-920-9	Sequence 9, Appli
23	30	90.4	2302	10	US-10-510-507-2	Sequence 2, Appli
24	30	90.4	2305	9	US-10-775-920-12	Sequence 12, Appli
25	30	90.4	2316	6	US-10-176-847-59	Sequence 59, Appli
26	30	90.4	2316	15	US-11-080-991-59	Sequence 59, Appli
27	30	90.4	2446	9	US-10-775-920-11	Sequence 11, Appli
28	30	90.4	2456	6	US-10-225-567A-453	Sequence 453, App
29	30	90.4	2456	7	US-10-269-909-63	Sequence 63, Appli
30	30	90.4	2456	7	US-10-269-909-64	Sequence 64, Appli
31	30	90.4	2456	7	US-10-295-027-619	Sequence 619, App
32	30	90.4	2456	8	US-10-600-816-2	Sequence 2, Appli
33	30	90.4	2456	8	US-10-600-816-18	Sequence 18, Appli
34	30	90.4	2456	9	US-10-775-920-10	Sequence 10, Appli
35	30	90.4	2456	10	US-10-936-626-40	Sequence 40, Appli
36	30	90.4	2456	10	US-10-938-061-40	Sequence 40, Appli
37	30	90.4	2456	16	US-11-169-041-32	Sequence 32, Appli
38	30	90.4	2593	7	US-10-264-049-834	Sequence 834, App
39	30	90.4	4239	6	US-10-198-846-10424	Sequence 10424, A
C 40	22.4	67.5	365	8	US-10-424-599-109942	Sequence 109942,
41	21	63.3	406	4	US-09-925-065A-342206	Sequence 342206,
42	21	63.3	406	4	US-09-925-065A-342207	Sequence 342207,
43	21	63.3	406	5	US-09-925-065A-342206	Sequence 342206,
44	21	63.3	406	5	US-09-925-065A-342207	Sequence 342207,
45	21	63.3	409	12	US-10-301-480-414987	Sequence 414987,
46	21	63.3	409	12	US-10-301-480-414988	Sequence 414988,
47	21	63.3	409	12	US-10-301-480-1028396	Sequence 1028396,
48	21	63.3	409	12	US-10-301-480-1028397	Sequence 1028397,
C 49	21	63.3	552	4	US-09-925-065A-407850	Sequence 407850,
C 50	21	63.3	552	5	US-09-925-065A-407850	Sequence 407850,
C 51	21	63.3	559	12	US-10-301-480-474416	Sequence 474416,
C 52	21	63.3	559	12	US-10-301-480-1087825	Sequence 1087825,
C 53	21	63.3	619	4	US-09-925-065A-42235	Sequence 42235, A
C 54	21	63.3	619	5	US-09-925-065A-42235	Sequence 42235, A
C 55	21	63.3	619	12	US-10-301-480-143473	Sequence 143473,
C 56	21	63.3	619	12	US-10-301-480-756882	Sequence 756882,
C 57	21	63.3	686	6	US-10-027-632-253967	Sequence 253967,
C 58	21	63.3	686	7	US-10-027-632-253967	Sequence 253967,
C 59	21	63.3	686	12	US-10-301-480-88757	Sequence 88757, A
C 60	21	63.3	686	12	US-10-301-480-702166	Sequence 702166,
C 61	21	63.3	698	4	US-09-925-065A-63800	Sequence 63800, A
C 62	21	63.3	698	5	US-09-925-065A-63800	Sequence 63800, A
C 63	21	63.3	698	12	US-10-301-480-185038	Sequence 185038,
C 64	21	63.3	698	12	US-10-301-480-78447	Sequence 78447,
C 65	21	63.3	710	8	US-10-767-701-19896	Sequence 19896, A
C 66	21	63.3	1101	11	US-10-932-182A-82765	Sequence 82765, A
C 67	20.8	62.7	2256646	8	US-10-470-565-1	Sequence 1, Appli
C 68	20.4	61.4	1074	3	US-09-738-626-595	Sequence 595, App
C 69	20.4	61.4	1201	9	US-10-494-672-81	Sequence 81, Appli
C 70	20.4	61.4	2292	11	US-10-932-182A-78725	Sequence 78725, A
C 71	20.4	61.4	2780	8	US-10-451-467A-363	Sequence 363, App
C 72	20.4	61.4	3309400	3	US-09-738-626-1	Sequence 1, Appli
C 73	20.2	60.8	201	9	US-10-719-993-31292	Sequence 31292, A
C 74	20.2	60.8	1146	8	US-10-282-122A-25373	Sequence 25373, A
C 75	20.2	60.8	775062	9	US-10-719-993-6844	Sequence 6844, App
C 76	20.2	60.8	783062	10	US-10-461-862-166	Sequence 166, App
C 77	20	60.2	282	9	US-10-674-124A-11800	Sequence 11800, A
C 78	20	60.2	381	8	US-10-437-963-66738	Sequence 66738, A
C 79	20	60.2	560	4	US-09-925-065A-219816	Sequence 219816,
C 80	20	60.2	560	4	US-09-925-065A-219817	Sequence 219817,
C 81	20	60.2	560	4	US-09-925-065A-219818	Sequence 219818,
C 82	20	60.2	560	5	US-09-925-065A-219816	Sequence 219816,
C 83	20	60.2	560	5	US-09-925-065A-219817	Sequence 219817,
C 84	20	60.2	560	5	US-09-925-065A-219818	Sequence 219818,
C 85	20	60.2	575	12	US-10-301-480-305477	Sequence 305477,
C 86	20	60.2	575	12	US-10-301-480-305478	Sequence 305478,
C 87	20	60.2	575	12	US-10-301-480-305479	Sequence 305479,
C 88	20	60.2	575	12	US-10-301-480-918886	Sequence 918886,
C 89	20	60.2	575	12	US-10-301-480-918887	Sequence 918887,
C 90	20	60.2	575	12	US-10-301-480-918888	Sequence 918888,















```
967 17.8 53.6 5833 7 US-10-275-140-3
968 17.8 53.6 5965 6 US-10-215-050-1
969 17.8 53.6 6028 7 US-10-215-050-3
970 17.8 53.6 6028 7 US-10-275-140-1
971 17.8 53.6 6028 9 US-10-723-860-2269
972 17.8 53.6 6028 10 US-10-367-057-137
973 17.8 53.6 6099 9 US-10-723-860-6484
974 17.8 53.6 6446 16 US-11-136-527-543
975 17.8 53.6 7760 13 US-11-097-143-21082
976 17.8 53.6 7947 13 US-11-097-143-39703
977 17.8 53.6 8091 7 US-10-101-510-86
978 17.8 53.6 8091 7 US-10-356-625-1
979 17.8 53.6 8091 7 US-10-159-563-322
980 17.8 53.6 8091 8 US-10-301-832-4
981 17.8 53.6 8091 10 US-10-631-467-364
982 17.8 53.6 13295 13 US-11-097-143-11473
983 17.8 53.6 13341 2 US-08-910-386A-1
984 17.8 53.6 15743 7 US-10-240-453-269
985 17.8 53.6 18753 13 US-11-097-143-28264
986 17.8 53.6 21354 8 US-10-221-714A-512
987 17.8 53.6 27425 8 US-10-773-714-18
988 17.8 53.6 40568 6 US-10-087-192-1573
989 17.8 53.6 44348 8 US-10-301-832-11
990 17.8 53.6 52764 6 US-10-087-192-127
991 17.8 53.6 70123 6 US-10-087-192-610
992 17.8 53.6 96594 3 US-09-997-722-22
993 17.8 53.6 110079 13 US-11-099-266-96
994 17.8 53.6 110079 13 US-11-099-266-96
995 17.8 53.6 160820 16 US-10-473-939-5
996 17.8 53.6 185750 6 US-11-114-798-56
997 17.8 53.6 196033 8 US-10-322-281-612
998 17.8 53.6 212231 6 US-10-087-192-1126
999 17.8 53.6 272032 11 US-10-330-773-102
c1000 17.8 53.6 304905 8 US-10-271-416-1
```

## ALIGNMENTS

```
RESULT 1
US-10-600-816-33
; Sequence 33, Application US/10600816
; Publication No. US20040121362A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: IDENTIFICATION AND MODULATION OF A G-PROTEIN COUPLED RECEPTOR
; FILE REFERENCE: D0251 NP
; CURRENT APPLICATION NUMBER: US/10/600,816
; CURRENT FILING DATE: 2003-06-20
; PRIOR APPLICATION NUMBER: U.S. 60/390,850
; PRIOR FILING DATE: 2002-06-20
; PRIOR APPLICATION NUMBER: U.S. 60/407,006
; PRIOR FILING DATE: 2002-08-29
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 33
; LENGTH: 42
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Degenerate Oligonucleotide.
; NAME/KEY: misc_feature
; LOCATION: (3)..(39)
; OTHER INFORMATION: wherein "n" equals A, C, G, or T.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6)..(33)
; OTHER INFORMATION: wherein "y" equals C, or T.
; FEATURE:
; NAME/KEY: misc_feature
```

```
; LOCATION: (27)..(42)
; OTHER INFORMATION: wherein "z" equals A, or G.
US-10-600-816-33
Query Match 100.0%; Score 33.2; DB 8; Length 42;
Best Local Similarity 100.0%; Pred. No. 6.5e-05;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCNCAYGCGTCCNTCNCNTAYAAARGAYTAYGARGTNAAR 42
Db 1 GCNCAYGCGTCCNTCNCNTAYAAARGAYTAYGARGTNAAR 42
RESULT 2
US-10-600-816-30
; Sequence 30, Application US/10600816
; Publication No. US20040121362A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: IDENTIFICATION AND MODULATION OF A G-PROTEIN COUPLED RECEPTOR
; FILE REFERENCE: D0251 NP
; CURRENT APPLICATION NUMBER: US/10/600,816
; CURRENT FILING DATE: 2003-06-20
; PRIOR APPLICATION NUMBER: U.S. 60/390,850
; PRIOR FILING DATE: 2002-06-20
; PRIOR APPLICATION NUMBER: U.S. 60/407,006
; PRIOR FILING DATE: 2002-08-29
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 30
; LENGTH: 42
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-600-816-30
Query Match 90.4%; Score 30; DB 8; Length 42;
Best Local Similarity 64.3%; Pred. No. 0.002;
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;
QY 1 GCNCAYGCGTCCNTCNCNTAYAAARGAYTAYGARGTNAAR 42
Db 1 GCCCAGCGTTGCCGAGCCCTTACAAGACTATGAAGTAAAG 42
RESULT 3
US-10-066-543-1937
; Sequence 1937, Application US/10066543
; Publication No. US20030087818A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Indrias, Carol Yoseph
; APPLICANT: Iodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Carter, Darick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Smith, Carole L.
; APPLICANT: Durham, Margarita
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.563
; CURRENT APPLICATION NUMBER: US/10/066,543
; CURRENT FILING DATE: 2002-01-31
; NUMBER OF SEQ ID NOS: 3417
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1937
; LENGTH: 497
; TYPE: DNA
; ORGANISM: Homo sapiens
```

## US-10-066-543-1937

Query Match 90.4%; Score 30; DB 6; Length 497;  
Best Local Similarity 64.3%; Pred. No. 0.004;  
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

Qy 1 GCNCAYGCGTGGCCNTCNCNTAYAAAGAYTAYGARGTNAAR 42  
|||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
Db 344 GCCCAGCGTTGGCGGAGCCCTTACAAAGACTATGAAGTAAAG 385

## RESULT 4

US-10-066-543-1811/c  
; Sequence 1811, Application US/10066543  
; Publication No. US20030087818A1  
; GENERAL INFORMATION:  
; APPLICANT: Jiang, Yugu  
; APPLICANT: Pyle, Ruth A.  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Indrias, Carol Yoseph  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Secrist, Heather  
; APPLICANT: Carter, Barrick  
; APPLICANT: Fanger, Gary R.  
; APPLICANT: Smith, Carole L.  
; APPLICANT: Durham, Margarita  
; APPLICANT: Stolk, John A.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; FILE OF INVENTION: AND DIAGNOSIS OF COLON CANCER  
; FILE REFERENCE: 210121.563  
; CURRENT APPLICATION NUMBER: US/10/066.543  
; CURRENT FILING DATE: 2002-01-31  
; NUMBER OF SEQ ID NOS: 3417  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1811  
; LENGTH: 552  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-066-543-1811

Query Match 90.4%; Score 30; DB 6; Length 552;  
Best Local Similarity 64.3%; Pred. No. 0.0041;  
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

Qy 1 GCNCAYGCGTGGCCNTCNCNTAYAAAGAYTAYGARGTNAAR 42  
|||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
Db 154 GCCCAGCGTTGGCGGAGCCCTTACAAAGACTATGAAGTAAAG 113

## RESULT 5

US-09-969-034-2222/c  
; Sequence 2222, Application US/09969034  
; Publication No. US20040110668A1  
; GENERAL INFORMATION:  
; APPLICANT: Burgess, Christopher C.  
; APPLICANT: Astle, Jon H.  
; APPLICANT: Carroll, Eddie III  
; APPLICANT: Catino, Theodore J.  
; APPLICANT: Dwivedi, Poornima  
; APPLICANT: Molino, Gary A.  
; APPLICANT: Thiagalingam, Arunthathi  
; APPLICANT: Lewis, Marcia E.  
; TITLE OF INVENTION: Nucleic Acid Sequences Differentially  
; FILE OF INVENTION: Expressed in Cancer Tissue  
; FILE REFERENCE: 1657/1032  
; CURRENT APPLICATION NUMBER: US/09/969.034  
; CURRENT FILING DATE: 2001-10-02  
; PRIOR FILING DATE: 60/237,271  
; PRIOR FILING DATE: 2000-02-10  
; NUMBER OF SEQ ID NOS: 4494  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2222  
; LENGTH: 620

## ; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: 491, 500, 517, 530, 545, 547, 553, 558, 570, 599, 620

; OTHER INFORMATION: n = A,T,C or G

US-09-969-034-2222

## Query Match

90.4%; Score 30; DB 3; Length 620;

Best Local Similarity 64.3%; Pred. No. 0.0043;

Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

Qy 1 GCNCAYGCGTGGCCNTCNCNTAYAAAGAYTAYGARGTNAAR 42

|||:||||| |||:||||| |||:||||| |||:||||| |||:|||||

Db 154 GCCCAGCGTTGGCGGAGCCCTTACAAAGACTATGAAGTAAAG 113

## RESULT 6

US-09-969-034-3393/c  
; Sequence 3393, Application US/09969034  
; Publication No. US20040110668A1  
; GENERAL INFORMATION:  
; APPLICANT: Burgess, Christopher C.  
; APPLICANT: Astle, Jon H.  
; APPLICANT: Carroll, Eddie III  
; APPLICANT: Catino, Theodore J.  
; APPLICANT: Dwivedi, Poornima  
; APPLICANT: Molino, Gary A.  
; APPLICANT: Thiagalingam, Arunthathi  
; APPLICANT: Lewis, Marcia E.  
; TITLE OF INVENTION: Nucleic Acid Sequences Differentially  
; FILE OF INVENTION: Expressed in Cancer Tissue  
; FILE REFERENCE: 1657/1032  
; CURRENT APPLICATION NUMBER: US/09/969.034  
; CURRENT FILING DATE: 2001-10-02  
; PRIOR FILING DATE: 60/237,271  
; PRIOR FILING DATE: 2000-02-10  
; NUMBER OF SEQ ID NOS: 4494  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3393  
; LENGTH: 634  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 363, 470, 472, 516, 523, 531, 551, 567, 573, 588, 594, 613,  
; LOCATION: 623  
; OTHER INFORMATION: n = A,T,C or G  
US-09-969-034-3393

## Query Match

90.4%; Score 30; DB 3; Length 634;

Best Local Similarity 64.3%; Pred. No. 0.0043;

Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

Qy 1 GCNCAYGCGTGGCCNTCNCNTAYAAAGAYTAYGARGTNAAR 42

|||:||||| |||:||||| |||:||||| |||:||||| |||:|||||

Db 158 GCCCAGCGTTGGCGGAGCCCTTACAAAGACTATGAAGTAAAG 117

## RESULT 7

US-10-125-968-701/c  
; Sequence 701, Application US/10125968  
; Publication No. US20030215805A1  
; GENERAL INFORMATION:  
; APPLICANT: Lillie, James  
; APPLICANT: Palermo, Adam  
; APPLICANT: Wang, Youzhen  
; APPLICANT: Steinmann, Kathleen  
; APPLICANT: Elias, Josh  
; APPLICANT: Mertens, Maureen  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND  
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, AND  
; TITLE OF INVENTION: THERAPY OF BREAST CANCER

```
; FILE REFERENCE: MRI-032
; CURRENT APPLICATION NUMBER: US/10/125,968
; CURRENT FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US 60/285,163
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 1417
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 701
; LENGTH: 642
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 72, 76, 77, 99, 101, 423, 445, 468, 489, 557, 566, 633
; OTHER INFORMATION: n = A,T,C or G
US-10-125-968-701

Query Match          90.4%; Score 30; DB 7; Length 642;
Best Local Similarity 64.3%; Pred. No. 0.0043; 8; Indels 0; Gaps 0;
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 GCNCAYGCGNTGGCCNTCNCNTAYAAAGAYTAYGARGTNAAR 42
    |||||
Db 184 GCCCAGCGTTGGCGGCGCCCTTACAAAGACTATGAAGTAAAG 143

RESULT 8
US-10-712-615-134
; Sequence 134, Application US/10712615
; Publication No. US20040214317A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: A NOVEL HUMAN G-PROTEIN COUPLED RECEPTOR, HGPBMY8, EXPRESSED
; FILE REFERENCE: D0047A-CIP
; CURRENT APPLICATION NUMBER: US/10/712,615
; CURRENT FILING DATE: 2003-11-13
; PRIOR APPLICATION NUMBER: U.S. 09/992,238
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: U.S. 60/248,285
; PRIOR FILING DATE: 2000-11-14
; PRIOR APPLICATION NUMBER: U.S. 60/268,581
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: U.S. 60/308,285
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: U.S. 60/317,166
; PRIOR FILING DATE: 2001-09-04
; NUMBER OF SEQ ID NOS: 134
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 134
; LENGTH: 1071
; TYPE: DNA
; ORGANISM: dna;Homo sapiens
US-10-712-615-134

Query Match          90.4%; Score 30; DB 9; Length 1071;
Best Local Similarity 64.3%; Pred. No. 0.005; 8; Indels 0; Gaps 0;
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 GCNCAYGCGNTGGCCNTCNCNTAYAAAGAYTAYGARGTNAAR 42
    |||||
Db 1018 GCCCAGCGTTGGCGGCGCCCTTACAAAGACTATGAAGTAAAG 1059

RESULT 9
US-09-866-050A-249
; Sequence 249, Application US/09866050A
; Publication No. US20030040471A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Orrust, Rene
```

```
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishanand D.
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; TITLE OF INVENTION: and Methods for Their Use
; FILE REFERENCE: 11000.1011c4U
; CURRENT APPLICATION NUMBER: US/09/866,050A
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 725
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 249
; LENGTH: 1212
; TYPE: DNA
; ORGANISM: Human
US-09-866-050A-249

Query Match          90.4%; Score 30; DB 3; Length 1212;
Best Local Similarity 64.3%; Pred. No. 0.0051; 8; Indels 0; Gaps 0;
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 GCNCAYGCGNTGGCCNTCNCNTAYAAAGAYTAYGARGTNAAR 42
    |||||
Db 990 GCCCAGCGTTGGCGGCGCCCTTACAAAGACTATGAAGTAAAG 1031

RESULT 10
US-10-152-661-249
; Sequence 249, Application US/10152661
; Publication No. US20030022835A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Orrust, Rene
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishanand D.
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; TITLE OF INVENTION: and Methods for Their Use
; FILE REFERENCE: 11000.1011c5
; CURRENT APPLICATION NUMBER: US/10/152,661
; CURRENT FILING DATE: 2002-05-20
; PRIOR APPLICATION NUMBER: 09/866,050
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/221,232
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: 60/206,650
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: 09/312,283
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: PCT/NZ99/00051
; PRIOR FILING DATE: 1999-04-29
; PRIOR APPLICATION NUMBER: 09/188,930
; PRIOR FILING DATE: 1998-11-09
; PRIOR APPLICATION NUMBER: 09/069,726
; PRIOR FILING DATE: 1998-04-29
; NUMBER OF SEQ ID NOS: 725
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 249
; LENGTH: 1212
; TYPE: DNA
; ORGANISM: Human
US-10-152-661-249

Query Match          90.4%; Score 30; DB 6; Length 1212;
Best Local Similarity 64.3%; Pred. No. 0.0051; 8; Indels 0; Gaps 0;
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 GCNCAYGCGNTGGCCNTCNCNTAYAAAGAYTAYGARGTNAAR 42
    |||||
Db 990 GCCCAGCGTTGGCGGCGCCCTTACAAAGACTATGAAGTAAAG 1031

RESULT 11
US-10-313-542-223
```

```
; Sequence 223, Application US/10313542
; Publication No. US20030120057A1
; GENERAL INFORMATION:
; APPLICANT: Roopa, Reddy
; APPLICANT: Guegler, Karl, J.
; APPLICANT: Au-Young, Janice
; TITLE OF INVENTION: COMPOSITION FOR DETECTION OF GENES ENCODING MEMBRANE-ASSOCIATED H
; FILE REFERENCE: PA-0013 US
; CURRENT APPLICATION NUMBER: US/10/313,542
; CURRENT FILING DATE: 2002-12-05
; PRIOR APPLICATION NUMBER: US/09/495,050
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/118,318
; PRIOR FILING DATE: 1999-02-01
; NUMBER OF SEQ ID NOS: 305
; SOFTWARE: PERL Program
; SEQ ID NO 223
; LENGTH: 1228
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030120057A1 2681738CT1
US-10-313-542-223

Query Match          90.4%; Score 30; DB 7; Length 1228;
Best Local Similarity 64.3%; Pred. No. 0.0052;
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 GCNCAYGCGTGGCCNTCNCNTAYAAAGAYTAYGARGTNAAR 42
   |||:||||| |||:||||| |||:||||| |||:|||||
Db 541 GCCACGCTTGGCCGAGCCCTTACAAGAGACTATGAAGTAAAG 582

RESULT 12
US-11-060-756-2418
; Sequence 2418, Application US/11060756
; Publication No. US20050221354A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William Martin
; TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
; FILE OF INVENTION: Target Genes
; FILE REFERENCE: AM101083 (031896-042000)
; CURRENT APPLICATION NUMBER: US/11/060,756
; CURRENT FILING DATE: 2005-02-18
; NUMBER OF SEQ ID NOS: 303284
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2418
; LENGTH: 1400
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-060-756-2418

Query Match          90.4%; Score 30; DB 13; Length 1400;
Best Local Similarity 64.3%; Pred. No. 0.0054;
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 GCNCAYGCGTGGCCNTCNCNTAYAAAGAYTAYGARGTNAAR 42
   |||:||||| |||:||||| |||:||||| |||:|||||
Db 215 GCCACGCTTGGCCGAGCCCTTACAAGAGACTATGAAGTAAAG 256

RESULT 13
US-11-060-756-6690
; Sequence 6690, Application US/11060756
; Publication No. US20050221354A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William Martin
; TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
; FILE OF INVENTION: Target Genes
; FILE REFERENCE: AM101083 (031896-042000)
```

```
; CURRENT APPLICATION NUMBER: US/11/060,756
; CURRENT FILING DATE: 2005-02-18
; NUMBER OF SEQ ID NOS: 303284
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6690
; LENGTH: 1400
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-060-756-6690

Query Match          90.4%; Score 30; DB 13; Length 1400;
Best Local Similarity 64.3%; Pred. No. 0.0054;
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 GCNCAYGCGTGGCCNTCNCNTAYAAAGAYTAYGARGTNAAR 42
   |||:||||| |||:||||| |||:||||| |||:|||||
Db 215 GCCACGCTTGGCCGAGCCCTTACAAGAGACTATGAAGTAAAG 256

RESULT 14
US-10-936-626-64
; Sequence 64, Application US/10936626
; Publication No. US20050106644A1
; GENERAL INFORMATION:
; APPLICANT: Cairns, Belinda
; APPLICANT: Chen, Ruihuan
; APPLICANT: Frantz, Gretchen
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Koepfen, Hartmut
; APPLICANT: Phillips, Heidi S.
; APPLICANT: Polakis, Paul
; APPLICANT: Spencer, Susan D.
; APPLICANT: Smith, Victoria
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wu, Thomas D.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: Compositions and Methods for the Diagnosis and
; FILE OF INVENTION: Treatment of Tumor
; FILE REFERENCE: P5001R1P1
; CURRENT APPLICATION NUMBER: US/10/936,626
; CURRENT FILING DATE: 2004-09-08
; PRIOR APPLICATION NUMBER: US 10/872,991
; PRIOR FILING DATE: 2004-06-21
; PRIOR APPLICATION NUMBER: US 10/872,972
; PRIOR FILING DATE: 2004-06-21
; PRIOR APPLICATION NUMBER: US 10/241,220
; PRIOR FILING DATE: 2002-09-11
; PRIOR APPLICATION NUMBER: US 10/177,488
; PRIOR FILING DATE: 2002-06-19
; PRIOR APPLICATION NUMBER: US 60/299,500
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 60/301,880
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/323,268
; PRIOR FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US 60/557,116
; PRIOR FILING DATE: 2004-03-26
; PRIOR APPLICATION NUMBER: US 60/598,899
; PRIOR FILING DATE: 2004-08-04
; NUMBER OF SEQ ID NOS: 154
; SEQ ID NO 64
; LENGTH: 1460
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-936-626-64

Query Match          90.4%; Score 30; DB 10; Length 1460;
Best Local Similarity 64.3%; Pred. No. 0.0054;
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 GCNCAYGCGTGGCCNTCNCNTAYAAAGAYTAYGARGTNAAR 42
   |||:||||| |||:||||| |||:||||| |||:|||||
Db 1145 GCCACGCTTGGCCGAGCCCTTACAAGAGACTATGAAGTAAAG 1186
```

```
RESULT 15
US-10-938-061-64
; Sequence 64, Application US/10938061
; Publication No. US20050107595A1
; GENERAL INFORMATION:
; APPLICANT: Cairns, Belinda
; APPLICANT: Chen, Ruihuan
; APPLICANT: Frantz, Gretchen
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Koepfen, Hartmut
; APPLICANT: Phillips, Heidi S.
; APPLICANT: Polakis, Paul
; APPLICANT: Spencer, Susan D.
; APPLICANT: Smith, Victoria
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wu, Thomas D.
; APPLICANT: Zhang, Zemin
; APPLICANT: Sakanaka, Chie
; APPLICANT: Chuntharapel, Anan
; APPLICANT: Reed Chae J.
; TITLE OF INVENTION: Compositions and Methods for the Diagnosis and
; FILE OF INVENTION: Treatment of Tumor
; FILE REFERENCE: P5001R1P1B
; CURRENT APPLICATION NUMBER: US/10/938,061
; CURRENT FILING DATE: 2004-09-10
; PRIOR APPLICATION NUMBER: US 10/872,991
; PRIOR FILING DATE: 2004-06-21
; PRIOR APPLICATION NUMBER: US 10/872,972
; PRIOR FILING DATE: 2004-06-21
; PRIOR APPLICATION NUMBER: US 10/241,220
; PRIOR FILING DATE: 2002-09-11
; PRIOR APPLICATION NUMBER: US 10/177,488
; PRIOR FILING DATE: 2002-06-19
; PRIOR APPLICATION NUMBER: US 60/299,500
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 60/301,880
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/323,268
; PRIOR FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US 60/557,116
; PRIOR FILING DATE: 2004-03-26
; PRIOR APPLICATION NUMBER: US 60/598,899
; PRIOR FILING DATE: 2004-08-04
; NUMBER OF SEQ ID NOS: 154
; SEQ ID NO 64
; LENGTH: 1460
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-938-061-64

Query Match          90.4%; Score 30; DB 10; Length 1460;
Best Local Similarity 64.3%; Pred. No. 0.0054;
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 GCNCAYGCGNTGCCNTCNCNTAYAAAGAYTAYGARGTNAAR 42
   |||:|||||  |||:|||||  |||:|||||  |||:|||||
Db 1145 GCCCAGCGTTGGCGGAGCCCTTACAAAGACTATGAAGTAAAG 1186

RESULT 16
US-10-224-289-5
; Sequence 5, Application US/10224289
; Publication No. US20030207288A1
; GENERAL INFORMATION:
; APPLICANT: LEWIN, DAVID A.
; APPLICANT: STEWART, TIMOTHY A.
; TITLE OF INVENTION: GPCR-LIKE RETINOIC ACID-INDUCED GENE 1 PROTEIN AND
; FILE OF INVENTION: NUCLEIC ACID
; FILE REFERENCE: 9800081-0085
; CURRENT APPLICATION NUMBER: US/10/224,289
; CURRENT FILING DATE: 2002-08-20
```

```
; PRIOR APPLICATION NUMBER: 60/313,940
; PRIOR FILING DATE: 2001-08-20
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1619
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-224-289-5

Query Match          90.4%; Score 30; DB 7; Length 1619;
Best Local Similarity 64.3%; Pred. No. 0.0056;
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 GCNCAYGCGNTGCCNTCNCNTAYAAAGAYTAYGARGTNAAR 42
   |||:|||||  |||:|||||  |||:|||||  |||:|||||
Db 1137 GCCCAGCGTTGGCGGAGCCCTTACAAAGACTATGAAGTAAAG 1178

RESULT 17
US-10-935-190-43
; Sequence 43, Application US/10935190
; Publication No. US20050037466A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: AU-YOUNG, Janice
; APPLICANT: BANDMAN, Olga
; APPLICANT: TANG, Y. Tom
; APPLICANT: YUE, Henry
; APPLICANT: AZIMZAI, Valda
; APPLICANT: BURFORD, Neil
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: LU, Dyung Aina M.
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: PATTERSON, Chandra
; APPLICANT: LAL, Preeti
; TITLE OF INVENTION: RECEPTORS AND ASSOCIATED PROTEINS
; FILE REFERENCE: PF-0726 PCT
; CURRENT APPLICATION NUMBER: US/10/935,190
; CURRENT FILING DATE: 2004-09-08
; PRIOR APPLICATION NUMBER: US/10/031,904
; PRIOR FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: 60/145,232; 60/158,578; 60/165,192
; PRIOR FILING DATE: 1999-07-21; 1999-10-07; 1999-11-12
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PERL Program
; SEQ ID NO 43
; LENGTH: 1619
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 2681738CB1
US-10-935-190-43

Query Match          90.4%; Score 30; DB 9; Length 1619;
Best Local Similarity 64.3%; Pred. No. 0.0056;
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 GCNCAYGCGNTGCCNTCNCNTAYAAAGAYTAYGARGTNAAR 42
   |||:|||||  |||:|||||  |||:|||||  |||:|||||
Db 1137 GCCCAGCGTTGGCGGAGCCCTTACAAAGACTATGAAGTAAAG 1178

RESULT 18
US-10-505-486-196
; Sequence 196, Application US/10505486
; Publication No. US20050118639A1
; GENERAL INFORMATION:
; APPLICANT: Takeda Chemical Industries, Ltd.
; TITLE OF INVENTION: Determination of a ligand
; FILE REFERENCE: P03-0006PCT
; CURRENT APPLICATION NUMBER: US/10/505,486
```



```
; CURRENT FILING DATE: 2004-08-20
; PRIOR APPLICATION NUMBER: JP 2002-45728
; PRIOR FILING DATE: 2002-02-22
; PRIOR APPLICATION NUMBER: JP 2002-213949
; PRIOR FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: JP 2002-298237
; PRIOR FILING DATE: 2002-10-11
; NUMBER OF SEQ ID NOS: 233
; SEQ ID NO 196
; LENGTH: 1788
; TYPE: DNA
; ORGANISM: Human
; US-10-505-486-196

Query Match          90.4%; Score 30; DB 10; Length 1788;
Best Local Similarity 64.3%; Pred. No. 0.0058;
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 GCNCAYGCNTGGCCNTCCNTAYAAARGAYTAYGARGTNAAR 42
Db 1018 GCCACGCTTGGCGGAGCCCTTACAAGACTATGAAGTAAG 1059

RESULT 19
US-10-775-920-13
; Sequence 13, Application US/10775920
; Publication No. US20040175744A1
; GENERAL INFORMATION:
; APPLICANT: Mergen Ltd
; TITLE OF INVENTION: POLYNUCLEOTIDE SEQUENCES AND CORRESPONDING ENCODED POLYPEPTIDES
; TITLE OF INVENTION: OF PARTICULAR SECRETED AND MEMBRANE-BOUND PROTEINS OVEREXPRESSED
; TITLE OF INVENTION: IN CERTAIN CANCERS
; FILE REFERENCE: Mergen - 0010B
; CURRENT APPLICATION NUMBER: US/10/775,920
; CURRENT FILING DATE: 2004-02-10
; PRIOR APPLICATION NUMBER: US 60/447,900
; PRIOR FILING DATE: 2003-02-13
; NUMBER OF SEQ ID NOS: 385
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 13
; LENGTH: 2297
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-775-920-13

Query Match          90.4%; Score 30; DB 9; Length 2297;
Best Local Similarity 64.3%; Pred. No. 0.0062;
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 GCNCAYGCNTGGCCNTCCNTAYAAARGAYTAYGARGTNAAR 42
Db 1125 GCCACGCTTGGCGGAGCCCTTACAAGACTATGAAGTAAG 1166

RESULT 20
US-10-224-289-3
; Sequence 3, Application US/10224289
; Publication No. US20030207288A1
; GENERAL INFORMATION:
; APPLICANT: LEWIN, DAVID A.
; APPLICANT: STEWART, TIMOTHY A.
; TITLE OF INVENTION: GPCR-LIKE RETINOIC ACID-INDUCED GENE 1 PROTEIN AND
; TITLE OF INVENTION: NUCLEIC ACID
; FILE REFERENCE: 980081-0085
; CURRENT APPLICATION NUMBER: US/10/224,289
; CURRENT FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: 60/313,940
; PRIOR FILING DATE: 2001-08-20
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 3
; LENGTH: 2302
; TYPE: DNA
```

```
; ORGANISM: Homo sapiens
; US-10-224-289-3

Query Match          90.4%; Score 30; DB 7; Length 2302;
Best Local Similarity 64.3%; Pred. No. 0.0062;
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 GCNCAYGCNTGGCCNTCCNTAYAAARGAYTAYGARGTNAAR 42
Db 1117 GCCACGCTTGGCGGAGCCCTTACAAGACTATGAAGTAAG 1158

RESULT 21
US-10-240-425-405
; Sequence 405, Application US/10240425
; Publication No. US20040033502A1
; GENERAL INFORMATION:
; APPLICANT: Williams, Amanda
; APPLICANT: Boland, Joseph F.
; APPLICANT: Lord, Reginald V.
; APPLICANT: Alvarez, Chris
; APPLICANT: Wetzel, Jon C.
; APPLICANT: Scherf, Uwe
; APPLICANT: Vockley, Joseph G.
; TITLE OF INVENTION: Gene Expression Profiles in Esophageal Tissue
; FILE REFERENCE: 44921-5026
; CURRENT APPLICATION NUMBER: US/10/240,425
; CURRENT FILING DATE: 2002-09-30
; PRIOR APPLICATION NUMBER: PCT/US01/09947
; PRIOR FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: US 60/193,446
; PRIOR FILING DATE: 2000-03-31
; NUMBER OF SEQ ID NOS: 1588
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 405
; LENGTH: 2302
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20040033502A1 AF095448
; US-10-240-425-405

Query Match          90.4%; Score 30; DB 8; Length 2302;
Best Local Similarity 64.3%; Pred. No. 0.0062;
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 GCNCAYGCNTGGCCNTCCNTAYAAARGAYTAYGARGTNAAR 42
Db 1117 GCCACGCTTGGCGGAGCCCTTACAAGACTATGAAGTAAG 1158

RESULT 22
US-10-775-920-9
; Sequence 9, Application US/10775920
; Publication No. US20040175744A1
; GENERAL INFORMATION:
; APPLICANT: Mergen Ltd
; TITLE OF INVENTION: POLYNUCLEOTIDE SEQUENCES AND CORRESPONDING ENCODED POLYPEPTIDES
; TITLE OF INVENTION: OF PARTICULAR SECRETED AND MEMBRANE-BOUND PROTEINS OVEREXPRESSED
; TITLE OF INVENTION: IN CERTAIN CANCERS
; FILE REFERENCE: Mergen - 0010B
; CURRENT APPLICATION NUMBER: US/10/775,920
; CURRENT FILING DATE: 2004-02-10
; PRIOR APPLICATION NUMBER: US 60/447,900
; PRIOR FILING DATE: 2003-02-13
; NUMBER OF SEQ ID NOS: 385
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9
; LENGTH: 2302
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-775-920-9
```

```
Query Match          90.4%; Score 30; DB 9; Length 2302;
Best Local Similarity 64.3%; Pred. No. 0.0062;
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 GCNCAYGNTGGCCNTCCNTAYAAAGAYTAYGARGTNAAR 42
Db 1117 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1158

RESULT 23
US-10-510-507-2
; Sequence 2, Application US/10510507
; Publication No. US20050282165A1
; GENERAL INFORMATION:
; APPLICANT: Terrett, Jonathan A
; TITLE OF INVENTION: DIAGNOSIS OF CARCINOMA USING RAIG1 POLYPEPTIDES
; FILE REFERENCE: 2543-1-039PCT/US
; CURRENT APPLICATION NUMBER: US/10/510,507
; CURRENT FILING DATE: 2004-10-07
; PRIOR APPLICATION NUMBER: GB0208331.9
; PRIOR FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: GB0221538.2
; PRIOR FILING DATE: 2002-09-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 2
; LENGTH: 2302
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-510-507-2

Query Match          90.4%; Score 30; DB 10; Length 2302;
Best Local Similarity 64.3%; Pred. No. 0.0062;
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 GCNCAYGNTGGCCNTCCNTAYAAAGAYTAYGARGTNAAR 42
Db 1117 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1158

RESULT 24
US-10-775-920-12
; Sequence 12, Application US/10775920
; Publication No. US20040175744A1
; GENERAL INFORMATION:
; APPLICANT: Mergen Ltd
; TITLE OF INVENTION: OF PARTICULAR SECRETED AND CORRESPONDING ENCODED POLYPEPTIDES
; TITLE OF INVENTION: OF PARTICULAR SECRETED AND MEMBRANE-BOUND PROTEINS OVEREXPRESSED
; FILE REFERENCE: Mergen - 0010B
; CURRENT APPLICATION NUMBER: US/10/775,920
; CURRENT FILING DATE: 2004-02-10
; PRIOR APPLICATION NUMBER: US 60/447,900
; PRIOR FILING DATE: 2003-02-13
; NUMBER OF SEQ ID NOS: 395
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 12
; LENGTH: 2305
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-775-920-12

Query Match          90.4%; Score 30; DB 9; Length 2305;
Best Local Similarity 64.3%; Pred. No. 0.0062;
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 GCNCAYGNTGGCCNTCCNTAYAAAGAYTAYGARGTNAAR 42
Db 1121 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1162

RESULT 25
US-10-176-847-59
```

```
; Sequence 59, Application US/10176847
; Publication No. US20030068636A1
; GENERAL INFORMATION:
; APPLICANT: Veiby, Petter Ole
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF BREAST
; FILE REFERENCE: MRI-039
; CURRENT APPLICATION NUMBER: US/10/176,847
; CURRENT FILING DATE: 2002-06-21
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 59
; LENGTH: 2316
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-176-847-59

Query Match          90.4%; Score 30; DB 6; Length 2316;
Best Local Similarity 64.3%; Pred. No. 0.0062;
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 GCNCAYGNTGGCCNTCCNTAYAAAGAYTAYGARGTNAAR 42
Db 1123 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1164

RESULT 26
US-11-080-991-59
; Sequence 59, Application US/11080991
; Publication No. US20050266437A1
; GENERAL INFORMATION:
; APPLICANT: Veiby, Petter Ole
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF BREAST
; TITLE OF INVENTION: AND OVARIAN CANCER
; FILE REFERENCE: MRI-039
; CURRENT APPLICATION NUMBER: US/11/080,991
; CURRENT FILING DATE: 2005-03-11
; PRIOR APPLICATION NUMBER: US/10/176,847
; PRIOR FILING DATE: 2002-06-21
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 59
; LENGTH: 2316
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-080-991-59

Query Match          90.4%; Score 30; DB 15; Length 2316;
Best Local Similarity 64.3%; Pred. No. 0.0062;
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 GCNCAYGNTGGCCNTCCNTAYAAAGAYTAYGARGTNAAR 42
Db 1123 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1164

RESULT 27
US-10-775-920-11
; Sequence 11, Application US/10775920
; Publication No. US20040175744A1
; GENERAL INFORMATION:
; APPLICANT: Mergen Ltd
; TITLE OF INVENTION: POLYNUCLEOTIDE SEQUENCES AND CORRESPONDING ENCODED POLYPEPTIDES
; TITLE OF INVENTION: OF PARTICULAR SECRETED AND MEMBRANE-BOUND PROTEINS OVEREXPRESSED
; FILE REFERENCE: Mergen - 0010B
; CURRENT APPLICATION NUMBER: US/10/775,920
; CURRENT FILING DATE: 2004-02-10
; PRIOR APPLICATION NUMBER: US 60/447,900
; PRIOR FILING DATE: 2003-02-13
; NUMBER OF SEQ ID NOS: 395
```

```
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11
; LENGTH: 2446
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-775-920-11

Query Match          90.4%; Score 30; DB 9; Length 2446;
Best Local Similarity 64.3%; Pred. No. 0.0063;
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 GCNCAYGCGTGGCCNTCNCNTAYAARGAYTAYGARGCTNAAR 42
Db 1271 GCCCAGCGTTGGCGGAGCCCTTACAAAGACTATGAAGTAAAG 1312

RESULT 28
US-10-225-567A-453
; Sequence 453, Application US/10225567A
; Publication No. US20030113798A1
; GENERAL INFORMATION:
; APPLICANT: LifeSpan Biosciences
; APPLICANT: Brown, Joseph P.
; APPLICANT: Burmer, Glenna C.
; APPLICANT: Roush, Christine L.
; TITLE OF INVENTION: ANTIGENIC PEPTIDES AND ANTIBODIES FOR G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 1920-4-4
; CURRENT APPLICATION NUMBER: US/10/225,567A
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/257,144
; PRIOR FILING DATE: 2000-12-19
; NUMBER OF SEQ ID NOS: 2292
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 453
; LENGTH: 2456
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-225-567A-453

Query Match          90.4%; Score 30; DB 6; Length 2456;
Best Local Similarity 64.3%; Pred. No. 0.0063;
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 GCNCAYGCGTGGCCNTCNCNTAYAARGAYTAYGARGCTNAAR 42
Db 1271 GCCCAGCGTTGGCGGAGCCCTTACAAAGACTATGAAGTAAAG 1312

RESULT 29
US-10-269-909-63
; Sequence 63, Application US/10269909
; Publication No. US20030180747A1
; GENERAL INFORMATION:
; APPLICANT: HRUBAN, RALPH H.
; APPLICANT: ARGANI, PEDRAM
; APPLICANT: IACOBUIZIO-DONAHUE, CHRISTINE
; APPLICANT: MAITRA, ANIRBAN
; TITLE OF INVENTION: PANCREATIC CANCER DIAGNOSIS AND THERAPIES
; FILE REFERENCE: 58303(71699)
; CURRENT APPLICATION NUMBER: US/10/269,909
; CURRENT FILING DATE: 2003-10-11
; PRIOR APPLICATION NUMBER: 60/328,609
; PRIOR FILING DATE: 2001-10-11
; PRIOR APPLICATION NUMBER: 60/332,754
; PRIOR FILING DATE: 2001-11-19
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 63
; LENGTH: 2456
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-269-909-63
```

```
Query Match          90.4%; Score 30; DB 7; Length 2456;
Best Local Similarity 64.3%; Pred. No. 0.0063;
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 GCNCAYGCGTGGCCNTCNCNTAYAARGAYTAYGARGCTNAAR 42
Db 1271 GCCCAGCGTTGGCGGAGCCCTTACAAAGACTATGAAGTAAAG 1312

RESULT 30
US-10-269-909-64
; Sequence 64, Application US/10269909
; Publication No. US20030180747A1
; GENERAL INFORMATION:
; APPLICANT: HRUBAN, RALPH H.
; APPLICANT: ARGANI, PEDRAM
; APPLICANT: IACOBUIZIO-DONAHUE, CHRISTINE
; APPLICANT: MAITRA, ANIRBAN
; TITLE OF INVENTION: PANCREATIC CANCER DIAGNOSIS AND THERAPIES
; FILE REFERENCE: 58303(71699)
; CURRENT APPLICATION NUMBER: US/10/269,909
; CURRENT FILING DATE: 2003-10-11
; PRIOR APPLICATION NUMBER: 60/328,609
; PRIOR FILING DATE: 2001-10-11
; PRIOR APPLICATION NUMBER: 60/332,754
; PRIOR FILING DATE: 2001-11-19
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 64
; LENGTH: 2456
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-269-909-64

Query Match          90.4%; Score 30; DB 7; Length 2456;
Best Local Similarity 64.3%; Pred. No. 0.0063;
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 GCNCAYGCGTGGCCNTCNCNTAYAARGAYTAYGARGCTNAAR 42
Db 1271 GCCCAGCGTTGGCGGAGCCCTTACAAAGACTATGAAGTAAAG 1312

RESULT 31
US-10-295-027-619
; Sequence 619, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
```

```
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 619
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-295-027-619

Query Match          90.4%; Score 30; DB 7; Length 2456;
Best Local Similarity 64.3%; Pred. No. 0.0063;
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY      1 GCNCAYGCGTGGCCNTCNCNTAYAAAGAYTAYGARGTNAAR 42
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1271 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1312

RESULT 32
US-10-600-816-2
; Sequence 2, Application US/10600816
; Publication No. US20040121362A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: IDENTIFICATION AND MODULATION OF A G-PROTEIN COUPLED RECEPTOR
; TITLE OF INVENTION: (GPCR), RAI3, ASSOCIATED WITH CHRONIC OBSTRUCTIVE PULMONARY
; TITLE OF INVENTION: DISEASE (COPD) AND NF-KB AND E-SELECTIN REGULATION
; FILE REFERENCE: D0251 NP
; CURRENT APPLICATION NUMBER: US/10/600,816
; CURRENT FILING DATE: 2003-06-20
; PRIOR APPLICATION NUMBER: U.S. 60/390,850
; PRIOR FILING DATE: 2002-06-20
; PRIOR APPLICATION NUMBER: U.S. 60/407,006
; PRIOR FILING DATE: 2002-08-29
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-600-816-2

Query Match          90.4%; Score 30; DB 8; Length 2456;
Best Local Similarity 64.3%; Pred. No. 0.0063;
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY      1 GCNCAYGCGTGGCCNTCNCNTAYAAAGAYTAYGARGTNAAR 42
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1271 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1312

RESULT 33
US-10-600-816-18
; Sequence 18, Application US/10600816
; Publication No. US20040121362A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: IDENTIFICATION AND MODULATION OF A G-PROTEIN COUPLED RECEPTOR
; TITLE OF INVENTION: (GPCR), RAI3, ASSOCIATED WITH CHRONIC OBSTRUCTIVE PULMONARY
; TITLE OF INVENTION: DISEASE (COPD) AND NF-KB AND E-SELECTIN REGULATION
; FILE REFERENCE: D0251 NP
; CURRENT APPLICATION NUMBER: US/10/600,816
; CURRENT FILING DATE: 2003-06-20
```

```
; PRIOR APPLICATION NUMBER: U.S. 60/390,850
; PRIOR FILING DATE: 2002-06-20
; PRIOR APPLICATION NUMBER: U.S. 60/407,006
; PRIOR FILING DATE: 2002-08-29
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 18
; LENGTH: 2456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: RAI3 Polymorphic Allele Summary Sequence.
; NAME/KEY: misc feature
; LOCATION: (112)..(112)
; OTHER INFORMATION: wherein "n" equals either G or A.
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (364)..(364)
; OTHER INFORMATION: wherein "n" equals either C or T.
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (511)..(511)
; OTHER INFORMATION: wherein "n" equals either C or T.
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (523)..(523)
; OTHER INFORMATION: wherein "n" equals either C or T.
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (605)..(605)
; OTHER INFORMATION: wherein "n" equals either A or G.
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (797)..(797)
; OTHER INFORMATION: wherein "n" equals either A or G.
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1111)..(1111)
; OTHER INFORMATION: wherein "n" equals either T or C.
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1173)..(1173)
; OTHER INFORMATION: wherein "n" equals either A or G.
; OTHER INFORMATION: wherein "n" equals either A or G.
US-10-600-816-18

Query Match          90.4%; Score 30; DB 8; Length 2456;
Best Local Similarity 64.3%; Pred. No. 0.0063;
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY      1 GCNCAYGCGTGGCCNTCNCNTAYAAAGAYTAYGARGTNAAR 42
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1271 GCCACGCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1312

RESULT 34
US-10-775-920-10
; Sequence 10, Application US/10775920
; Publication No. US20040175744A1
; GENERAL INFORMATION:
; APPLICANT: Mergen Ltd
; TITLE OF INVENTION: POLYNUCLEOTIDE SEQUENCES AND CORRESPONDING ENCODED POLYPEPTIDES
; TITLE OF INVENTION: OF PARTICULAR SECRETED AND MEMBRANE-BOUND PROTEINS OVEREXPRESSED
; TITLE OF INVENTION: IN CERTAIN CANCERS
; FILE REFERENCE: Mergen - 0010B
; CURRENT APPLICATION NUMBER: US/10/775,920
; CURRENT FILING DATE: 2004-02-10
; PRIOR APPLICATION NUMBER: US 60/447,900
; PRIOR FILING DATE: 2003-02-13
; NUMBER OF SEQ ID NOS: 395
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 10
; LENGTH: 2456
```

; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-775-920-10

Query Match 90.4%; Score 30; DB 9; Length 2456;  
Best Local Similarity 64.3%; Pred. No. 0.0063;  
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

Qy 1 GCNCAYCGTGGCCNTCCNTAAYARGAYTAYGAGTNAAR 42  
Db 1271 GCCCAGCCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1312

## RESULT 35

US-10-936-626-40  
; Sequence 40, Application US/10936626

; Publication No. US20050106644A1

; GENERAL INFORMATION:

; APPLICANT: Cairns, Belinda

; APPLICANT: Chen, Ruihuan

; APPLICANT: Frantz, Gretchen

; APPLICANT: Hillan, Kenneth J.

; APPLICANT: Koepfen, Hartmut

; APPLICANT: Phillips, Heidi S.

; APPLICANT: Polakis, Paul

; APPLICANT: Spencer, Susan D.

; APPLICANT: Smith, Victoria

; APPLICANT: Williams, P. Mickey

; APPLICANT: Wu, Thomas D.

; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: Compositions and Methods for the Diagnosis and

; FILE REFERENCE: P5001R1P1

; CURRENT APPLICATION NUMBER: US/10/936,626

; PRIOR FILING DATE: 2004-09-08

; PRIOR APPLICATION NUMBER: US 10/872,991

; PRIOR FILING DATE: 2004-06-21

; PRIOR APPLICATION NUMBER: US 10/872,972

; PRIOR FILING DATE: 2004-06-21

; PRIOR APPLICATION NUMBER: US 10/241,220

; PRIOR FILING DATE: 2002-09-11

; PRIOR APPLICATION NUMBER: US 10/177,488

; PRIOR FILING DATE: 2002-06-19

; PRIOR APPLICATION NUMBER: US 60/299,500

; PRIOR FILING DATE: 2001-06-20

; PRIOR APPLICATION NUMBER: US 60/301,880

; PRIOR FILING DATE: 2001-06-29

; PRIOR APPLICATION NUMBER: US 60/323,268

; PRIOR FILING DATE: 2001-09-18

; PRIOR APPLICATION NUMBER: US 60/557,116

; PRIOR FILING DATE: 2004-03-26

; PRIOR APPLICATION NUMBER: US 60/598,899

; PRIOR FILING DATE: 2004-08-04

; NUMBER OF SEQ ID NOS: 154

; SEQ ID NO 40

; LENGTH: 2456

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-936-626-40

Query Match 90.4%; Score 30; DB 10; Length 2456;  
Best Local Similarity 64.3%; Pred. No. 0.0063;  
Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

Qy 1 GCNCAYCGTGGCCNTCCNTAAYARGAYTAYGAGTNAAR 42  
Db 1271 GCCCAGCCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1312

## RESULT 36

US-10-938-061-40

; Sequence 40, Application US/10938061

; Publication No. US20050107595A1

; GENERAL INFORMATION:

; APPLICANT: Cairns, Belinda

; APPLICANT: Chen, Ruihuan

; APPLICANT: Frantz, Gretchen

; APPLICANT: Hillan, Kenneth J.

; APPLICANT: Koepfen, Hartmut

; APPLICANT: Phillips, Heidi S.

; APPLICANT: Polakis, Paul

; APPLICANT: Spencer, Susan D.

; APPLICANT: Smith, Victoria

; APPLICANT: Williams, P. Mickey

; APPLICANT: Wu, Thomas D.

; APPLICANT: Zhang, Zemin

; APPLICANT: Sakanaka, Chie

; APPLICANT: Chuntharapai, Anan

; APPLICANT: Reed Chae J.

; TITLE OF INVENTION: Compositions and Methods for the Diagnosis and

; FILE REFERENCE: P5001R1P1B

; CURRENT APPLICATION NUMBER: US/10/938,061

; PRIOR FILING DATE: 2004-09-10

; PRIOR APPLICATION NUMBER: US 10/872,991

; PRIOR FILING DATE: 2004-06-21

; PRIOR APPLICATION NUMBER: US 10/872,972

; PRIOR FILING DATE: 2004-06-21

; PRIOR APPLICATION NUMBER: US 10/241,220

; PRIOR FILING DATE: 2002-09-11

; PRIOR APPLICATION NUMBER: US 10/177,488

; PRIOR FILING DATE: 2002-06-19

; PRIOR APPLICATION NUMBER: US 60/299,500

; PRIOR FILING DATE: 2001-06-20

; PRIOR APPLICATION NUMBER: US 60/301,880

; PRIOR FILING DATE: 2001-06-29

; PRIOR APPLICATION NUMBER: US 60/323,268

; PRIOR FILING DATE: 2001-09-18

; PRIOR APPLICATION NUMBER: US 60/557,116

; PRIOR FILING DATE: 2004-03-26

; PRIOR APPLICATION NUMBER: US 60/598,899

; PRIOR FILING DATE: 2004-08-04

; NUMBER OF SEQ ID NOS: 154

; SEQ ID NO 40

; LENGTH: 2456

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-938-061-40

Query Match 90.4%; Score 30; DB 10; Length 2456;

Best Local Similarity 64.3%; Pred. No. 0.0063;

Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

Qy 1 GCNCAYCGTGGCCNTCCNTAAYARGAYTAYGAGTNAAR 42  
Db 1271 GCCCAGCCTTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1312

## RESULT 37

US-11-169-041-32

; Sequence 32, Application US/11169041

; Publication No. US20060019284A1

; GENERAL INFORMATION:

; APPLICANT: Bristol-Myers Squibb Company

; TITLE OF INVENTION: IDENTIFICATION OF POLYNUCLEOTIDES FOR PREDICTING ACTIVITY OF

; TITLE OF INVENTION: COMPOUNDS THAT INTERACT WITH AND/OR MODULATE PROTEIN TYROSINE

; TITLE OF INVENTION: KINASES AND/OR PROTEIN TYROSINE KINASE PATHWAYS IN LUNG CANCER

; TITLE OF INVENTION: CELLS

; FILE REFERENCE: 10001 NP

; CURRENT APPLICATION NUMBER: US/11/169,041

; PRIOR FILING DATE: 2005-06-28

; PRIOR APPLICATION NUMBER: 60/584,405

; PRIOR FILING DATE: 2004-06-30

; NUMBER OF SEQ ID NOS: 527

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 32

; LENGTH: 2456

; TYPE: DNA

; ORGANISM: Homo sapiens

US-11-169-041-32

Query Match 90.4%; Score 30; DB 16; Length 2456;

Best Local Similarity 64.3%; Pred. No. 0.0063; 8; Indels 0; Gaps 0;

Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 GCNCAYGCGTGGCCNTCNCNTAYAAAGAYTAYGAGTNAAR 42

DB 1271 GCCCAGCGCTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1312

RESULT 38

US-10-264-049-834

; Sequence 834, Application US/10264049

; Publication No. US20040005579A1

; GENERAL INFORMATION:

; APPLICANT: Birse et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: P133P1

; CURRENT APPLICATION NUMBER: US/10/264, 049

; CURRENT FILING DATE: 2002-10-04

; PRIOR APPLICATION NUMBER: PCT/US01/18569

; PRIOR FILING DATE: 2001-06-07

; PRIOR APPLICATION NUMBER: US 60/209,467

; PRIOR FILING DATE: 2000-06-07

; NUMBER OF SEQ ID NOS: 4360

; SOFTWARE: PatentIn Ver. 3.1

; SEQ ID NO 834

; LENGTH: 2593

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-264-049-834

Query Match 90.4%; Score 30; DB 7; Length 2593;

Best Local Similarity 64.3%; Pred. No. 0.0064; 8; Indels 0; Gaps 0;

Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 GCNCAYGCGTGGCCNTCNCNTAYAAAGAYTAYGAGTNAAR 42

DB 1383 GCCCAGCGCTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1424

RESULT 39

US-10-198-846-10424

; Sequence 10424, Application US/10198846

; Publication No. US20030099974A1

; GENERAL INFORMATION:

; APPLICANT: Lillie, James

; APPLICANT: Xu, Yongyao

; APPLICANT: Wang, Youzhen

; APPLICANT: Steinmann, Kathleen

; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS

; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND

; FILE REFERENCE: MRI-049

; CURRENT APPLICATION NUMBER: US/10/198,846

; CURRENT FILING DATE: 2002-07-18

; PRIOR APPLICATION NUMBER: 60/306,220

; PRIOR FILING DATE: 2001-07-18

; NUMBER OF SEQ ID NOS: 14084

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 10424

; LENGTH: 4239

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: 1, 4238, 4239

; OTHER INFORMATION: n = A,T,C or G

US-10-198-846-10424

Query Match 90.4%; Score 30; DB 6; Length 4239;

Best Local Similarity 64.3%; Pred. No. 0.0074; 8; Indels 0; Gaps 0;

Matches 27; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 GCNCAYGCGTGGCCNTCNCNTAYAAAGAYTAYGAGTNAAR 42

DB 1679 GCCCAGCGCTGGCCGAGCCCTTACAAAGACTATGAAGTAAAG 1720

RESULT 40

US-10-424-599-109942/c

; Sequence 109942, Application US/10424599

; Publication No. US20040031072A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J

; APPLICANT: Kovalic David K

; APPLICANT: Zhou Yihua

; APPLICANT: Cao Yongwei

; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With

; FILE REFERENCE: 38-21(53223)B

; CURRENT APPLICATION NUMBER: US/10/424,599

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 285684

; SEQ ID NO 109942

; LENGTH: 365

; TYPE: DNA

; ORGANISM: Glycine max

; FEATURE:

; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_70290C.1

US-10-424-599-109942

Query Match 67.5%; Score 22.4; DB 8; Length 365;

Best Local Similarity 59.0%; Pred. No. 12;

Matches 23; Conservative 5; Mismatches 11; Indels 0; Gaps 0;

QY 4 CAYGCGTGGCCNTCNCNTAYAAAGAYTAYGAGTNAAR 42

DB 181 CAGGAATGGCCCTCTCCCATAAAGATGATGACAAG 143

RESULT 41

US-09-925-065A-342206

; Sequence 342206, Application US/09925065A

; Publication No. US20040181048A1

; GENERAL INFORMATION:

; APPLICANT: Wang, David G.

; TITLE OF INVENTION: Identification and Mapping of Single

; FILE REFERENCE: 108827.135

; CURRENT APPLICATION NUMBER: US/09/925,065A

; CURRENT FILING DATE: 2001-08-08

; PRIOR APPLICATION NUMBER: US 60/243,096

; PRIOR FILING DATE: 2000-10-24

; PRIOR APPLICATION NUMBER: US 60/252,147

; PRIOR FILING DATE: 2000-11-20

; PRIOR APPLICATION NUMBER: US 60/250,092

; PRIOR FILING DATE: 2000-11-30

; PRIOR APPLICATION NUMBER: US 60/261,766

; PRIOR FILING DATE: 2001-01-16

; PRIOR APPLICATION NUMBER: US 60/289,846

; PRIOR FILING DATE: 2001-05-09

; NUMBER OF SEQ ID NOS: 957086

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 342206

; LENGTH: 406

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-925-065A-342206

Query Match 63.3%; Score 21; DB 4; Length 406;

Best Local Similarity 51.2%; Pred. No. 56;







```
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 407850
; LENGTH: 552
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-407850
```

```
Query Match 63.3%; Score 21; DB 5; Length 552;
Best Local Similarity 51.2%; Pred. No. 61;
Matches 21; Conservative 7; Mismatches 13; Indels 0; Gaps 0;

Qy 2 CNCAYGCGTGGCCNTCCNTAYAAAGAYTAYGARGTNAAR 42
Db 308 CTCTTTCTCTCTCTCAATATGAAGATTATGAGGACAA 268
```

```
RESULT 51
US-10-301-480-474416/c
; Sequence 474416, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 474416
; LENGTH: 559
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-474416
```

```
Query Match 63.3%; Score 21; DB 12; Length 559;
Best Local Similarity 51.2%; Pred. No. 62;
Matches 21; Conservative 7; Mismatches 13; Indels 0; Gaps 0;

Qy 2 CNCAYGCGTGGCCNTCCNTAYAAAGAYTAYGARGTNAAR 42
Db 315 CTCTTTCTCTCTCTCAATATGAAGATTATGAGGACAA 275
```

```
RESULT 52
US-10-301-480-1087825/c
; Sequence 1087825, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
```

```
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1087825
; LENGTH: 559
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-1087825
```

```
Query Match 63.3%; Score 21; DB 12; Length 559;
Best Local Similarity 51.2%; Pred. No. 62;
Matches 21; Conservative 7; Mismatches 13; Indels 0; Gaps 0;

Qy 2 CNCAYGCGTGGCCNTCCNTAYAAAGAYTAYGARGTNAAR 42
Db 315 CTCTTTCTCTCTCTCAATATGAAGATTATGAGGACAA 275
```

```
RESULT 53
US-09-925-065A-42235/c
; Sequence 42235, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 42235
; LENGTH: 619
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-42235
```

```
Query Match 63.3%; Score 21; DB 4; Length 619;
Best Local Similarity 55.3%; Pred. No. 63;
Matches 21; Conservative 6; Mismatches 11; Indels 0; Gaps 0;

Qy 1 GNCAYGCGTGGCCNTCCNTAYAAAGAYTAYGARGT 38
Db 39 GCCATTTCATGGACATCTCCACACAAAGTCTACGGACT 2
```

```
RESULT 54
US-09-925-065A-42235/c
; Sequence 42235, Application US/09925065A
; Publication No. US2005028172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
```

; PRIOR FILING DATE: 2002-08

; CURRENT FILING DATE: 2002-04-30  
 ; PRIOR APPLICATION NUMBER: IIS 60/218 006



```
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 63800
; LENGTH: 698
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-63800

Query Match      63.3%; Score 21; DB 5; Length 698;
Best Local Similarity 51.2%; Pred. No. 66;
Matches 21; Conservative 7; Mismatches 13; Indels 0; Gaps 0;

QY      2  CNCAYCNCNTGCCNTCCNTAARGAYTAYGARGTNAAR 42
Db      235  CTCTTCTCAGCCTTCCCAATGTAAGACGATGAGGTAAA 275

RESULT 63
US-10-301-480-165038
; Sequence 165038, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms
; FILE OF INVENTION: in the Human Genome
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 165038
; LENGTH: 698
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-165038

Query Match      63.3%; Score 21; DB 12; Length 698;
Best Local Similarity 51.2%; Pred. No. 66;
Matches 21; Conservative 7; Mismatches 13; Indels 0; Gaps 0;

QY      2  CNCAYCNCNTGCCNTCCNTAARGAYTAYGARGTNAAR 42
Db      235  CTCTTCTCAGCCTTCCCAATGTAAGACGATGAGGTAAA 275

RESULT 64
US-10-301-480-778447
; Sequence 778447, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms
; FILE OF INVENTION: in the Human Genome
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 778447
```

```
; LENGTH: 698
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-778447

Query Match      63.3%; Score 21; DB 12; Length 698;
Best Local Similarity 51.2%; Pred. No. 66;
Matches 21; Conservative 7; Mismatches 13; Indels 0; Gaps 0;

QY      2  CNCAYCNCNTGCCNTCCNTAARGAYTAYGARGTNAAR 42
Db      235  CTCTTCTCAGCCTTCCCAATGTAAGACGATGAGGTAAA 275

RESULT 65
US-10-767-701-19896/c
; Sequence 19896, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 19896
; LENGTH: 710
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(710)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB5049-011-R1-XP1-A5
US-10-767-701-19896

Query Match      63.3%; Score 21; DB 8; Length 710;
Best Local Similarity 58.3%; Pred. No. 66;
Matches 21; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

QY      2  CNCAYCNCNTGCCNTCCNTAARGAYTAYGARG 37
Db      131  CGCCCGGTAGCGTCCGCCGACAGGACGACG 96

RESULT 66
US-10-932-182A-82765
; Sequence 82765, Application US/10932182A
; Publication No. US20060046253A1
; GENERAL INFORMATION:
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHIISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: 030685-043
; CURRENT APPLICATION NUMBER: US/10/932,182A
; CURRENT FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 82765
; LENGTH: 1101
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-10-932-182A-82765

Query Match      63.3%; Score 21; DB 11; Length 1101;
Best Local Similarity 55.3%; Pred. No. 75;
```



RESULT 71  
US-10-451-467A-363  
; Sequence 363, Application US/10451467A  
; Publication No. US20040161840A1  
; GENERAL INFORMATION:  
; APPLICANT: CONTRERAS, ROLAND HENRI  
; APPLICANT: EBERHARDT, INES  
; APPLICANT: LUYTEN, WALTER HERMAN MARIA LOUIS  
; APPLICANT: REEKWANS, RIEKA JOSEPHINA  
; TITLE OF INVENTION: BAX-RESPONSIVE GENES FOR DRUG TARGET IDENTIFICATION IN  
; TITLE OF INVENTION: YEAST AND FUNGI  
; FILE REFERENCE: JAB-1667  
; CURRENT APPLICATION NUMBER: US/10/451,467A  
; CURRENT FILING DATE: 2003-06-19  
; PRIOR APPLICATION NUMBER: EP 00870318.3  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: EP 01870002.1  
; PRIOR FILING DATE: 2001-01-04  
; PRIOR APPLICATION NUMBER: EP 01870003.9  
; PRIOR FILING DATE: 2001-01-09  
; NUMBER OF SEQ ID NOS: 732  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 363  
; LENGTH: 2780  
; TYPE: DNA  
; ORGANISM: Saccharomyces cerevisiae  
US-10-451-467A-363

Query Match 61.4%; Score 20.4; DB 8; Length 2780;  
Best Local Similarity 56.8%; Pred. No. 1.8e+02;  
Matches 21; Conservative 5; Mismatches 11; Indels 0; Gaps 0;  
QY 6 YGCGTGGCCNTCCNTAYAAAGAYTAYGAGTNAAR 42  
DB 2716 CGCGTGGCCATCACCATACATTGGTTAGGAGGAAAA 2752

RESULT 72  
US-09-738-626-1/c  
; Sequence 1, Application US/09738625  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 1  
; LENGTH: 3309400  
; TYPE: DNA  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-1

Query Match 61.4%; Score 20.4; DB 3; Length 3309400;

Best Local Similarity 52.5%; Pred. No. 1.2e+03;  
Matches 21; Conservative 6; Mismatches 13; Indels 0; Gaps 0;  
QY 1 GCNCAYGCGTGGCCNTCCNTAYAAAGAYTAYGAGTNA 40  
DB 552828 GCGCAGCGCTTGACCACGCGCTTGTGAGAGACGACCAAGTGA 552789  
RESULT 73  
US-10-719-993-31292/c  
; Sequence 31292, Application US/10719993  
; Publication No. US20040265849A1  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele et al.  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001496  
; CURRENT APPLICATION NUMBER: US/10/719,993  
; CURRENT FILING DATE: 2003-11-24  
; NUMBER OF SEQ ID NOS: 55342  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 31292  
; LENGTH: 201  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-719-993-31292

Query Match 60.8%; Score 20.2; DB 9; Length 201;  
Best Local Similarity 56.4%; Pred. No. 1.1e+02;  
Matches 22; Conservative 4; Mismatches 13; Indels 0; Gaps 0;  
QY 2 CMCAYGCGTGGCCNTCCNTAYAAAGAYTAYGAGTNA 40  
DB 88 CACACGCTTCCCATCTCCTTTCAATGAATGTGTGTCA 50

RESULT 74  
US-10-282-122A-25373/c  
; Sequence 25373, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22



```
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; TITLE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 66738
; LENGTH: 381
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_67661C.1
US-10-437-963-66738

Query Match          60.2%; Score 20; DB 8; Length 381;
Best Local Similarity 54.1%; Pred. No. 1.8e+02;
Matches 20; Conservative 6; Mismatches 11; Indels 0; Gaps 0;

QY      6 YGCGTGGCCNTCCNTAYAAARGAYTAYGARGTNAAR 42
      :|||:|||:|||:|||:|||:|||:|||:|||:|||:
Db      105 TGCATGACCTCAACAGATATAATAGTACGAGTAGAA 69

RESULT 79
US-09-925-065A-219816/c
; Sequence 219816, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925, 065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 219816
; LENGTH: 560
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-219816

Query Match          60.2%; Score 20; DB 4; Length 560;
Best Local Similarity 51.3%; Pred. No. 1.8e+02;
Matches 20; Conservative 7; Mismatches 12; Indels 0; Gaps 0;

QY      4 CAYGCGTGGCCNTCCNTAYAAARGAYTAYGARGTNAAR 42
      ||:|||:|||:|||:|||:|||:|||:|||:|||:
Db      270 CATGCGTGGCCACCACCAATATCAACATTTTAAATGAAA 232

RESULT 80
US-09-925-065A-219817/c
; Sequence 219817, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
```

```
; CURRENT APPLICATION NUMBER: US/09/925, 065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 219817
; LENGTH: 560
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-219817

Query Match          60.2%; Score 20; DB 4; Length 560;
Best Local Similarity 51.3%; Pred. No. 1.8e+02;
Matches 20; Conservative 7; Mismatches 12; Indels 0; Gaps 0;

QY      4 CAYGCGTGGCCNTCCNTAYAAARGAYTAYGARGTNAAR 42
      ||:|||:|||:|||:|||:|||:|||:|||:|||:
Db      270 CATGCGTGGCCACCACCAATATCAACATTTTAAATGAAA 232

RESULT 81
US-09-925-065A-219818/c
; Sequence 219818, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925, 065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 219818
; LENGTH: 560
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-219818

Query Match          60.2%; Score 20; DB 4; Length 560;
Best Local Similarity 51.3%; Pred. No. 1.8e+02;
Matches 20; Conservative 7; Mismatches 12; Indels 0; Gaps 0;

QY      4 CAYGCGTGGCCNTCCNTAYAAARGAYTAYGARGTNAAR 42
      ||:|||:|||:|||:|||:|||:|||:|||:|||:
Db      270 CATGCGTGGCCACCACCAATATCAACATTTTAAATGAAA 232

RESULT 82
US-09-925-065A-219816/c
; Sequence 219816, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
```



```

; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 219816
; LENGTH: 560
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-925-065A-219816

```

Query Match 60.2%; Score 20; DB 5; Length 560;  
Best Local Similarity 51.3%; Pred. No. 1.8e+02;  
Matches 20; Conservative 7; Mismatches 12; Indels 0; Gaps 0;

**Qy**

4 CAYGNTGGCCNCTNCNCNTAYAARGAYTAYGARGTNAAR 42  
||:||:||||| ||:||: ||:||: ||:||:

**Dp**

270 CATGCCTGGCCACCAACAATATCAACATTTTTAAATGAAA 232

RESULT 83

```

US-09-925-065A-219817/C
; Sequence 219817, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 219817
; LENGTH: 560
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-219817

```

```
Query Match      60.2%; Score 20; DB 5; Length 560;
Best Local Similarity 51.3%; Pred. No. 1.8e+02;
Matches 20: Conservative 7; Mismatches 12; Indels 0; Gaps 0;
```

QY		4	CAYGCNTGGCCNCTNCNCNTAYAARGAYTAYGARGTNAAR	42
			:                         :     :	
Dd		270	CATGCCTGGGCCCAACAATAATCAACATTTCATTTAAATGAAG	232

RESULT 84

RESOLUTION 84  
US-09-925-065A-219818/c  
; Sequence 219818, Application US/09925065A

```

; Publication No. US20050228172A9
;
; GENERAL INFORMATION:
;
; APPLICANT: Wang, David G.
;
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
;
; FILE REFERENCE: 108827.135
;
; CURRENT APPLICATION NUMBER: US/09/925,065A
;
; CURRENT FILING DATE: 2001-08-08
;
; PRIOR APPLICATION NUMBER: US 60/243,096
;
; PRIOR FILING DATE: 2000-10-24
;
; PRIOR APPLICATION NUMBER: US 60/252,147
;
; PRIOR FILING DATE: 2000-11-20
;
; PRIOR APPLICATION NUMBER: US 60/250,092
;
; PRIOR FILING DATE: 2000-11-30
;
; PRIOR APPLICATION NUMBER: US 60/261,766
;
; PRIOR FILING DATE: 2001-01-16
;
; PRIOR APPLICATION NUMBER: US 60/289,846
;
; PRIOR FILING DATE: 2001-05-09
;
; NUMBER OF SEQ ID NOS: 957086
;
; SOFTWARE: Fast-SEQ for Windows Version 4.0
;
; SEQ ID NO 219818
;
; LENGTH: 560
;
; TYPE: DNA
;
; ORGANISM: Homo sapiens
;
; US-09-925-065A-219818

```

Query Match	60.2%	Score 20;	DB 5;	Length 560;
Best Local Similarity	51.3%	Pred. No. 1.8e+02;		
Matches	20:	Conservative	7:	Mismatches 12:
				Indels 0:
				Gaps 0:

**Qy**

4 CAYGCNTGGCCNCTNCNCNTAYAARGAYTAYGARGTNAAR 42  
||:||| ||||| | | ||: | : | : | : | : | :

**Db**

270 CATGCCTGGCCACCACAATATCAACATTCTTTTAATAATGAAGA 232

## RESULT 85

```

US-10-301-480-305477/c
; Sequence 305477, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms
; TITLE OF INVENTION: in the Human Genome
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1236818
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 305477
; LENGTH: 575
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-305477

```

Query Match	Score 20;	DB 12;	Length 575;
Best Local Similarity	51.3%	Pred. No. 1.8e+02;	
Matches 20: Conservative	7;	Mismatches 12;	Indels 0;
Matches 20: Conservative	7;	Mismatches 12;	Indels 0;
Matches 20: Conservative	7;	Mismatches 12;	Indels 0;

QY 4 CAYGNTGGCCNTCNCNTAYAARGAYTAYGAGTNAAR 42  
||:||||||| | ||: |: |: |: |: |: ||:  
nb 270 CATGCCTGGCCACCACAATATCAACATTTTTAATGA AA 232

98 T. J. Prescott

US-10-301-480-305478/c  
; Sequence 305478, Application US/10301480  
; Publication No. US20060057564A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.



Best Local Similarity 51.3%; Pred. No. 1.8e+02;  
Matches 20; Conservative 7; Mismatches 12; Indels 0; Gaps 0;

**Qy**            4 CAYGCNTGGCCNTNCNCNTAYAARGAYTAYGARGTNAAR 42  
||: || ||||| | ||: ||: ||: ||: ||:  
**Db**         270 CATGCCTGGCCACCAATATCAACATTTTTAANTGAAA 232

```

RESULT 91
US-10-312-273-340/c
, Sequence 340, Application US/10312273
, Publication No. US20040005667A1
, GENERAL INFORMATION:
, APPLICANT: CHIRON SPA
, TITLE OF INVENTION: IMMUNISATION AGAINST CHLAMYDIA PNEUMONIAE
, FILE REFERENCE: P025035WO
, CURRENT APPLICATION NUMBER: US/10/312,273
, CURRENT FILING DATE: 2002-12-20
, PRIOR APPLICATION NUMBER: 0016363.4
, PRIOR FILING DATE: 2000-07-03
, PRIOR APPLICATION NUMBER: 0017047.2
, PRIOR FILING DATE: 2000-07-11
, PRIOR APPLICATION NUMBER: 0017983.8
, PRIOR FILING DATE: 2000-07-21
, PRIOR APPLICATION NUMBER: 0019368.0
, PRIOR FILING DATE: 2000-08-07
, PRIOR APPLICATION NUMBER: 0020440.4
, PRIOR FILING DATE: 2000-08-18
, PRIOR APPLICATION NUMBER: 0022593.9
, PRIOR FILING DATE: 2000-09-14
, PRIOR APPLICATION NUMBER: 0027549.5
, PRIOR FILING DATE: 2000-11-10
, PRIOR APPLICATION NUMBER: 0031706.5
, PRIOR FILING DATE: 2000-12-22
, NUMBER OF SEQ ID NOS: 664
, SOFTWARE: SeqWin99, version 1.02
, SEQ ID NO 340
, LENGTH: 669
, TYPE: DNA
, ORGANISM: Chlamydia pneumoniae
US-10-312-273-340

```

Query Match	60.2%	Score 20;	DB 7;	Length 689;
Best Local Similarity	57.1%	Pred. No. 1.9e+02;		
Matches 20;	Conservative	5;	Mismatches 10;	Indels 0;
				Gaps 0;

**Qy** 2 C NCAYGCTGGCCNTCNCCTAYAAAGATYAYGAR 36  
| | | | | | | | | | | | | | : | : | :  
**Db** 513 CTCATGCATAGGCTTTACCATAGAAGATAACGAG 479

```

RESULT 92
US-09-764-869-1560/c
; Sequence 1560, Application US/09764869
; Patent No. US20020061521A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC007
; CURRENT APPLICATION NUMBER: US/09/764,869
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2442
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1560
; LENGTH: 5161
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-869-1560

```

```
Query Match      60.2%; Score 20; DB 3; Length 5161;
Best Local Similarity 62.5%; Pred. No. 3.4e+02;
Matches 20; Conservative 4; Mismatches 8; Indels 0; Gaps 0;
```

Qy 4 CAYGNTGGCCNTCNCNTAYAAAGAYTAYGA 35  
||:||||||| | :||:|  
Db 4936 CATGCTGGCCGTCATATCTTAATGATTATGA 4905

```

RESULT 93
US-10-091-504-1560/c
; Sequence 1560, Application US/10091504
; Publication No. US20030059908A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC007C1
; CURRENT APPLICATION NUMBER: US/10/091,504
; CURRENT FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 2442
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1560
; LENGTH: 5161
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-091-504-1560

```

```
Query Match      60.2%; Score 20; DB 6; Length 5161;
Best Local Similarity 62.5%; Pred. No. 3.4e+02;
Matches 20; Conservative 4; Mismatches 8; Indels 0; Gaps 0;
```

Qy 4 CAYGCTGGCCNTCCNTAYAAAGTAYTA 35  
|||:|||||:|:|:|:|:  
Db 4936 CATGCTGGCCGTCATATCTAATGATTATGA 4905

```

RESULT 94
US-10-227-577-1560/c
; Sequence 1560, Application US/10227577
; Publication No. US20040005575A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC007C2
; CURRENT APPLICATION NUMBER: US/10/227,577
; CURRENT FILING DATE: 2002-08-26
; PRIOR APPLICATION NUMBER: 10/091,504
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 09/764,869
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 2442
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1560
; LENGTH: 5161
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-227-577-1560

```

```
Query Match          60.2%; Score 20; DB 7; Length 5161;
Best Local Similarity 62.5%; Pred. No. 3.4e+02;
Matches 20; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY      4 CAYGCTGCGCCNTCCNCNTAYAAAGAYTAYGA 35
      ||:|||||:|||||:|||||:|||||:|||||:
Db      4936 CATGCTGCGCCGCATATTCCTATGATTATGA 4905

RESULT 95
US-10-289-762-1
; Sequence 1, Application US/10289762
; Publication No. US20040006218A1
; GENERAL INFORMATION:
; APPLICANT: Grifflais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/10/289,762
; CURRENT FILING DATE: 2003-03-27
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 1
; LENGTH: 1230025
; TYPE: DNA
; ORGANISM: Chlamydia pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(15000)
; OTHER INFORMATION: n=a or c or g or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (15001)..(30000)
; OTHER INFORMATION: n=a or c or g or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (30001)..(45000)
; OTHER INFORMATION: n=a or c or g or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (45001)..(60000)
; OTHER INFORMATION: n=a or c or g or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (60001)..(75000)
; OTHER INFORMATION: n=a or c or g or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (75001)..(90000)
; OTHER INFORMATION: n=a or c or g or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (90001)..(105000)
; OTHER INFORMATION: n=a or c or g or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (105001)..(120000)
; OTHER INFORMATION: n=a or c or g or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (120001)..(135000)
; OTHER INFORMATION: n=a or c or g or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (135001)..(150000)
; OTHER INFORMATION: n=a or c or g or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (150001)..(165000)
; OTHER INFORMATION: n=a or c or g or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (165001)..(180000)
; OTHER INFORMATION: n=a or c or g or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (180001)..(195000)
; OTHER INFORMATION: n=a or c or g or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (195001)..(210000)
; OTHER INFORMATION: n=a or c or g or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (210001)..(225000)
; OTHER INFORMATION: n=a or c or g or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (225001)..(240000)
; OTHER INFORMATION: n=a or c or g or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (240001)..(255000)
; OTHER INFORMATION: n=a or c or g or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (255001)..(270000)
; OTHER INFORMATION: n=a or c or g or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (270001)..(285000)
; OTHER INFORMATION: n=a or c or g or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (285001)..(300000)
; OTHER INFORMATION: n=a or c or g or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (30001)..(315000)
; OTHER INFORMATION: n=a or c or g or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (315001)..(330000)
; OTHER INFORMATION: n=a or c or g or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (330001)..(345000)
; OTHER INFORMATION: n=a or c or g or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (345001)..(360000)
; OTHER INFORMATION: n=a or c or g or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (360001)..(375000)
; OTHER INFORMATION: n=a or c or g or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (375001)..(390000)
; OTHER INFORMATION: n=a or c or g or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (390001)..(405000)
; OTHER INFORMATION: n=a or c or g or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (405001)..(420000)
; OTHER INFORMATION: n=a or c or g or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (420001)..(435000)
; OTHER INFORMATION: n=a or c or g or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (435001)..(450000)
; OTHER INFORMATION: n=a or c or g or t
```

```
FEATURE:
NAME/KEY: misc feature
LOCATION: (450001)..(465000)
OTHER INFORMATION: n=a or c or g or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (465001)..(480000)
OTHER INFORMATION: n=a or c or g or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (480001)..(495000)
OTHER INFORMATION: n=a or c or g or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (495001)..(510000)
OTHER INFORMATION: n=a or c or g or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (510001)..(525000)
OTHER INFORMATION: n=a or c or g or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (525001)..(540000)
OTHER INFORMATION: n=a or c or g or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (540001)..(555000)
OTHER INFORMATION: n=a or c or g or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (555001)..(570000)
OTHER INFORMATION: n=a or c or g or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (570001)..(585000)
OTHER INFORMATION: n=a or c or g or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (585001)..(600000)
OTHER INFORMATION: n=a or c or g or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (600001)..(615000)
OTHER INFORMATION: n=a or c or g or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (615001)..(630000)
OTHER INFORMATION: n=a or c or g or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (630001)..(645000)
OTHER INFORMATION: n=a or c or g or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (645001)..(660000)
OTHER INFORMATION: n=a or c or g or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (660001)..(675000)
OTHER INFORMATION: n=a or c or g or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (675001)..(690000)
OTHER INFORMATION: n=a or c or g or t
FEATURE:
```

```
Query Match 60.2%; Score 20; DB 7; Length 1230025;
Best Local Similarity 57.1%; Pred. No. 1.5e+03;
Matches 20; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

QY 2 CCAAGTGGTCCNTCCNTAARGAYTAYGAR 36
Db 835280 CTCATGATAGGCTTTACCATAGAAAGATAACGAG 835314
```

```
RESULT 96
US-09-925-065A-519993/c
; Sequence 519993, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 519993
; LENGTH: 570
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-519993
```

```
Query Match 59.6%; Score 19.8; DB 4; Length 570;
Best Local Similarity 55.3%; Pred. No. 2.2e+02;
Matches 21; Conservative 5; Mismatches 12; Indels 0; Gaps 0;

QY 5 AYCNTGGCCNTCCNTAARGAYTAYGARGTNAAR 42
Db 454 ATCGATGGCATTAAAGGTTAAAGATTAGGAACCTTAAG 417
```

```
RESULT 97
US-09-925-065A-519993/c
; Sequence 519993, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 519993
; LENGTH: 570
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-519993
```

```
Query Match 59.6%; Score 19.8; DB 5; Length 570;
Best Local Similarity 55.3%; Pred. No. 2.2e+02;
Matches 21; Conservative 5; Mismatches 12; Indels 0; Gaps 0;
```

